

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

ÉVALUATION DU NIVEAU D'ACQUISITION DES EXPRESSIONS FIGÉES CHEZ
DES ENFANTS ALLOPHONES ET FRANCOPHONES DE LA 3^E À LA 6^E ANNÉE
DU PRIMAIRE

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE EN LINGUISTIQUE

PAR

JAËLLE ARCHAMBAULT-LAPOINTE

JUILLET 2009

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.01-2006). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

Les têtes se forment sur les langages, les pensées prennent la teinte des idiomes, l'esprit, en chaque langue, a sa forme particulière.

[Jean-Jacques Rousseau]

Pour les locutions, il était insatiable de renseignements, car, leur supposant parfois un sens plus précis qu'elles n'ont, il eût désiré savoir ce qu'ont voulait dire exactement par celles qu'il entendait le plus souvent employer : la beauté du diable, du sang bleu, une vie de bâton de chaise, le quart d'heure de Rabelais, être le prince des élégances, donner carte blanche, être réduit à quia, etc., et dans quels cas déterminés il pouvait à son tour les faire figurer dans ses propos. À leur défaut, il plaçait des jeux de mots qu'il avait appris.

[Marcel Proust]

REMERCIEMENTS

Je ne saurais accepter le mérite de ce mémoire sans manifester ma sincère reconnaissance à tous ceux et celles qui ont contribué de près ou de loin à ce projet.

J'aimerais d'abord adresser mes remerciements à ma directrice de mémoire, Lori Morris, pour son enthousiasme et son intérêt envers mon travail, son dévouement dans l'organisation de la collecte de données, ainsi que ses conseils judicieux à chaque étape de la rédaction de mon mémoire. Soulignons également la contribution de mes lecteurs, Tom Cobb et Gladys Jean, qui ont généreusement accepté d'évaluer ce mémoire.

J'aimerais par ailleurs exprimer ma reconnaissance à la direction, aux enseignants, aux parents des élèves et aux élèves de l'école primaire où s'est déroulée la collecte de données, pour avoir accepté de bon gré de participer à ma recherche.

Dans un autre ordre d'idées, je me dois d'exprimer ma gratitude à l'égard du Conseil de recherche en sciences humaines du Canada (CRSH), qui grâce à son soutien financier, a su créer des conditions favorables à la poursuite de mes études de maîtrise.

À titre personnel, je souhaite remercier chaleureusement les membres de ma famille proche pour m'avoir épaulée et guidée tout au long de mon cheminement scolaire, ainsi que mon conjoint, pour son attitude positive, sa patience et son appui indéfectible pendant toutes mes études universitaires. Merci également à tous mes amis et collègues qui se sont intéressés à mon projet et qui m'ont encouragée à poursuivre.

À vous tous, bonne lecture !

TABLE DES MATIÈRES

LISTE DES TABLEAUX	vii
LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES	viii
RÉSUMÉ	ix
INTRODUCTION	1
CHAPITRE I	
PROBLÉMATIQUE	5
1.1 Les expressions figées dans l'acquisition des langues et dans la communication.....	5
1.2 Statut particulier des élèves allophones de Montréal	8
1.3 Élèves allophones à Montréal : survol d'une démographie scolaire	9
1.4 Développement des connaissances linguistiques chez les allophones et les francophones	11
1.5 Conclusion du chapitre	13
CHAPITRE II	
CADRE THÉORIQUE	14
2.1 Tour d'horizon du domaine de la phraséologie : quelques repères terminologiques pour s'y retrouver.....	15
2.1.1 Collocations.....	15
2.1.2 Séquence préfabriquée.....	16
2.1.3 Expressions idiomatiques	18
2.1.4 Locutions	20
2.1.5 Expression figée.....	21
2.1.6 Vers une définition des expressions figées	22

2.2	Caractéristiques des expressions figées	24
2.2.1	Caractéristiques cognitives	24
2.2.2	Caractéristiques liées à l'usage	25
2.2.3	Caractéristiques syntaxiques	28
2.2.4	Caractéristiques lexicales	29
2.2.5	Caractéristiques sémantiques	30
2.2.6	Caractéristiques phonologiques	33
2.2.7	Le figement : un phénomène relatif	34
2.3	La place du langage figé dans l'acquisition des langues	35
2.3.1	Le langage figé et l'acquisition de la langue première	36
2.3.2	Le langage figé et l'acquisition d'une langue seconde	40
2.3.3	L'emploi d'expressions figées pour atteindre les compétences d'un locuteur natif	46
2.4	Conclusion du chapitre	47
CHAPITRE III		
	QUESTIONS DE RECHERCHE	49
CHAPITRE IV		
	APPROCHE MÉTHODOLOGIQUE	53
4.1	Participants	54
4.2	Choix de l'instrument de mesure	57
4.3	Choix des livres pour l'extraction des expressions figées	58
4.4	Description des expressions entrant dans la composition du test	59
4.5	Extraction des expressions figées	60
4.6	Sélection des expressions figées	61
4.7	Conception de l'instrument de mesure	62
4.7.1	Coupure des mots	63
4.7.2	Adaptations effectuées en fonction des participants	65
4.7.3	Questionnaires complémentaires	67
4.8	Validation du test	68
4.8.1	Échantillon	68
4.8.2	Milieu scolaire	69
4.9	Collecte de données	69

4.10	Compilation des résultats	70
4.11	Le calcul du score	72
4.12	Analyses	73
4.13	Conclusion du chapitre	73
CHAPITRE V		
	ANALYSE STATISTIQUE ET INTERPRÉTATION DES RÉSULTATS	75
5.1	Analyse par année scolaire	76
5.2	Analyse par groupe linguistique	78
5.3	Analyse par sexe	80
5.4	Analyse de la fréquence des expressions, de la fréquence des mots clés et du nombre de concurrents	81
5.5	Analyse par item	85
5.5.1	Performance des élèves aux items	85
5.5.2	L'effet du facteur « année scolaire » sur la réussite des items	92
5.5.3	Items reflétant une influence du facteur « groupe linguistique »	95
5.6	Conclusion	96
	CONCLUSION	99
	BIBLIOGRAPHIE	105
	ANNEXES	116

LISTE DES TABLEAUX

TABLEAU 4.1 RÉPARTITION DES PARTICIPANTS SELON L'ANNÉE, LE GROUPE LINGUISTIQUE ET LE SEX.....	55
TABLEAU 4.2 ÂGE DES PARTICIPANTS SELON L'ANNÉE SCOLAIRE ET LE GROUPE LINGUISTIQUE.....	55
TABLEAU 5.1 TAUX DE SUCCÈS PAR ANNÉE SCOLAIRE	76
TABLEAU 5.2 TAUX DE RÉUSSITE PAR ANNÉE ET PAR GROUPE LINGUISTIQUE.....	78
TABLEAU 5.3 TAUX DE RÉUSSITE PAR ANNÉE ET PAR SEXE	80
TABLEAU 5.4 CARACTÉRISTIQUES DES ITEMS : FRÉQUENCE DES EF, FRÉQUENCE DES MOTS CIBLÉS ET NOMBRE DE CONCURRENTS DES MOTS CIBLÉS.	82
TABLEAU 5.5 TAUX DE RÉUSSITE DE CHAQUE ITEM SELON L'ANNÉE SCOLAIRE	86

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

BNC	British National Corpus
CANCODE	Cambridge and Nottingham Corpus of Discourse in English
CELE	Centre for English Language Education
CSDM	Commission scolaire de Montréal
EF	Expression figée
GDT	Grand dictionnaire terminologique
LADL	Laboratoire d'analyse documentaire et linguistique
L1	Langue première ou langue maternelle
L2	Langue seconde
LN	Locuteur natif
LNN	Locuteur non natif
MELS	Ministère de l'Éducation, du Loisir et du Sport
MEQ	Ministère de l'Éducation
OQLF	Office québécois de la langue française

RÉSUMÉ

Cette recherche exploratoire avait pour objectif de mesurer l'influence de divers facteurs individuels et linguistiques sur la connaissance d'élèves de niveau primaire en matière d'expressions figées françaises. Les facteurs individuels que nous avons étudiés sont le niveau scolaire, le groupe linguistique et le sexe. Quant aux facteurs linguistiques, il s'agissait de la fréquence des expressions, de la fréquence des mots clés et du nombre de concurrents plus fréquents que les mots clés¹. Notre population était formée d'élèves allophones et francophones de la 3^e à la 6^e année fréquentant une école francophone de Montréal. Les groupes linguistiques ont été déterminés en fonction de la langue parlée à la maison (avec les parents).

Les élèves (165 au total) ont été évalués dans leur milieu scolaire, au moyen d'un test de production écrite guidée. Celui-ci était présenté sous la forme de phrases lacunaires comprenant des expressions figées dont un ou deux mots étaient tronqués. La tâche des participants consistait à compléter les expressions. À partir des données recueillies, nous avons effectué des analyses statistiques afin de déterminer si les facteurs considérés avaient eu une influence significative sur la performance des élèves. Ces analyses ont permis de constater que la réussite des élèves au test était fortement liée au niveau scolaire. En effet, les élèves plus avancés ont mieux performé que les élèves moins avancés, et cet écart s'est révélé significatif. Ce résultat suggère que plus les élèves avancent dans leur scolarisation, plus ils développent leurs connaissances des expressions. Par contre, nos analyses n'ont pas révélé de différence significative entre les groupes linguistiques (allophones et francophones). Parmi les explications possibles, nous avons évoqué le fait que la grande majorité des enfants allophones fréquentaient un établissement scolaire francophone depuis la maternelle ; il est également possible que la complexité des situations linguistiques dans les familles des enfants allophones entre en jeu. Quant au facteur « sexe », nous n'avons pas observé d'écart de performance entre les garçons et les filles. En ce qui a trait aux caractéristiques linguistiques, la fréquence des mots clés ainsi que le nombre de concurrents plus fréquents ont eu un effet significatif sur la performance des élèves. Par contre, nos

¹ Les « mots clés » correspondent aux mots tronqués contenus dans les expressions du test de production guidée. Le nombre de « concurrents » fait quant à lui référence au nombre de mots plus fréquents commençant par la ou les mêmes lettres que les mots tronqués.

analyses n'ont pas révélé de lien entre la fréquence des expressions et la réussite des enfants. Il est probable que la base de données utilisée pour mesurer la fréquence soit en cause.

Mots clés : acquisition du français langue seconde, expressions figées, connaissances lexicales, élèves du primaire, allophones, francophones.

INTRODUCTION

La maîtrise d'une langue repose sur un ensemble complexe de connaissances et de compétences. Pour atteindre un haut niveau de compétence dans une langue donnée, il ne suffit pas de connaître un grand nombre de mots et de savoir appliquer les règles de grammaire : il est également essentiel de comprendre et d'utiliser les expressions propres à cette langue, car celles-ci permettent de s'exprimer de façon juste et naturelle dans la langue cible. En effet, les adultes locuteurs natifs préfèrent généralement employer des expressions courantes plutôt que de produire de nouvelles combinaisons lexicales. Par exemple, si un locuteur natif du français veut signifier qu'une personne s'est enfuie très rapidement, il aura tendance à employer des expressions familières comme « elle s'est enfuie à toutes jambes » ou « elle est partie en coup de vent », plutôt que d'avoir recours à d'autres tournures moins courantes, comme « elle a fui très vite ». De même, s'il veut décrire l'intensité d'une difficulté, il sera davantage porté à employer l'expression « de peine et de misère » au lieu de « avec beaucoup de difficulté ». Ces exemples montrent que pour bien s'exprimer dans une langue, on doit produire un discours qui non seulement est orthographiquement et grammaticalement correct, mais qui privilégie les formules naturelles dans la langue cible.

Les expressions d'usage forment une proportion très importante de la langue parlée et écrite : le français en compterait plusieurs dizaines de milliers (Mel'čuk 1993). Parmi les termes les plus fréquemment employés dans la recherche anglo-saxonne pour désigner les différents types d'expressions, on trouve *formulaic*

language, formulas, prefabricated language, ready-made language, chunks, unanalysed language et *wholes* (Weinert 1995).² La terminologie foisonne également dans la recherche française : *collocations, locutions, locutions figées, expressions idiomatiques, expressions figées, stéréotypes*, sont quelques-uns des termes les plus souvent employés (Forsberg 2006). Bien que ces appellations désignent des phénomènes linguistiques apparentés, ces derniers comportent toutefois des nuances, notamment en ce qui a trait à leur structure grammaticale et à leur contenu sémantique (nous développerons ce point dans le *Cadre théorique*). Par conséquent, le choix d'un terme et d'une définition appropriés se révèle souvent une tâche délicate. Précisons d'emblée que dans le cadre de la recherche présentée ici, c'est le terme « expression figée »³, appellation empruntée à Svensson (2004), qui a été adopté pour désigner les expressions conventionnelles que partagent les locuteurs du français.

Tel que mentionné précédemment, le langage figé est un phénomène très répandu dans l'usage. Puisque les locuteurs possèdent la capacité de combiner les mots pour créer de nouveaux énoncés, il y a lieu de se demander pourquoi les formules « toutes faites » occupent une place si importante dans la langue. Plusieurs raisons d'ordre cognitif et socio-interactionnel sont susceptibles d'expliquer ce phénomène. L'une des raisons fondamentales en est que l'emploi d'expressions figées constitue un véritable outil de communication pour les locuteurs de tous les âges et de tous les niveaux de compétence dans la langue cible. Chez les jeunes enfants d'âge préscolaire natifs et non natifs, l'utilisation de phrases-clés préfabriquées permettrait de communiquer plus efficacement (Wong-Fillmore 1976, Peters 1983, Wray 2002). Chez les locuteurs adultes, l'emploi d'EF contribuerait de

² Wray (2002) a recensé plus d'une cinquantaine de termes anglophones associés au langage figé (voir annexe).

³ Nous définirons ce terme plus en détail dans le chapitre intitulé *Cadre théorique*. Il est à noter qu'afin d'alléger le texte, nous emploierons également l'abréviation « EF » pour désigner les expressions figées. Nous emploierons également le terme « langage figé » lorsque nous parlerons des expressions de façon plus générale.

façon significative à la fluidité verbale (Wood 2006; Pawley et Syder 2000) et à l'idiomaticité (Wray 2002)⁴.

L'emploi d'EF est donc d'une grande utilité pour les locuteurs natifs et non natifs, et ce, à bien des égards. Or, pour être à même de tirer profit des EF dans la communication, il faut évidemment être capable de les reconnaître et de les employer. Cependant, il y a lieu de croire que les locuteurs moins expérimentés, tels que les enfants et les locuteurs non natifs, peuvent éprouver des difficultés lorsqu'il s'agit de reconnaître, de comprendre et d'utiliser correctement des expressions figées. En effet, des études ayant examiné les facteurs menant à l'acquisition des expressions figées suggèrent que l'âge des enfants jouerait un rôle dans la capacité de compréhension et de produire des expressions, et en particulier lorsqu'il s'agit d'expressions peu fréquentes ou sémantiquement opaques (par ex. Nippold et Rudzinski 1993; Nippold et Taylor 1995). D'autres études révèlent que les locuteurs natifs et non natifs auraient des approches différentes par rapport à l'apprentissage de la langue cible, et qu'ils n'auraient pas nécessairement la même représentation mentale du lexique (Wray 2002). Puisque de nombreux facteurs individuels et linguistiques semblent avoir un effet sur l'acquisition des expressions figées, il nous paraît important et intéressant de savoir dans quelle mesure ces facteurs influencent le niveau de connaissance des EF dans un contexte d'apprentissage précis, à savoir une école primaire francophone où se côtoient des enfants locuteurs natifs du français (francophones) et locuteurs non natifs du français (allophones). C'est pourquoi nous nous pencherons sur ces enjeux dans la présente étude.

Objectifs et orientation de la recherche

L'objectif de cette recherche est de connaître l'influence possible de divers facteurs individuels et linguistiques sur les connaissances d'enfants allophones et francophones de la 3^e à la 6^e année du primaire en matière d'expressions figées.

⁴ Les fonctions du langage figé feront l'objet d'une discussion plus développée dans la *Problématique* et dans le *Cadre théorique*.

Dans cette visée, nous allons évaluer des élèves au moyen d'un test de production guidée. La tâche consistera à compléter des expressions figées : elle sollicitera donc à la fois les connaissances réceptives et productives des enfants. À partir des données recueillies, nous procéderons à des analyses statistiques afin de savoir si la performance des élèves semble être influencée par les facteurs individuels suivants : l'année scolaire, le groupe linguistique (allophones et francophones) et le sexe. Nous tenterons également de déterminer si certaines caractéristiques linguistiques (la fréquence des expressions, la fréquence des mots clés et le nombre de concurrents) semblent avoir un effet sur la performance des élèves (nous expliquerons les caractéristiques linguistiques ultérieurement). Les objectifs de cette étude ayant été présentés, il convient maintenant d'indiquer l'orientation adoptée. Mentionnons d'abord qu'il s'agit d'une recherche transversale de type évaluative : nous évaluerons tous les groupes de participants simultanément. Soulignons également que cette étude se veut de nature exploratoire : à notre connaissance, les recherches concernant l'acquisition des expressions en français chez les enfants allophones et francophones de Montréal sont très rares. Ajoutons par ailleurs que beaucoup d'études ont porté sur les EF en anglais langue seconde, alors que très peu d'études ont porté sur l'acquisition d'EF par des apprenants du français langue seconde (Forsberg 2006). La présente recherche constituera donc un premier pas dans cette direction.

CHAPITRE I

PROBLÉMATIQUE

Avant de présenter le cadre théorique de cette recherche, il convient de démontrer au lecteur la pertinence d'une étude sur les expressions figées en acquisition des langues secondes. Dans cette optique, nous décrirons brièvement les fonctions qu'occupent les EF dans l'acquisition d'une langue première et d'une langue seconde, ainsi que dans la communication orale et écrite, tant pour les locuteurs natifs que non natifs. Par la suite, nous expliquerons en quoi la présente étude s'insère dans son contexte social et pédagogique. Afin de mieux décrire la situation démographique des allophones, nous donnerons également quelques statistiques révélatrices sur leurs effectifs dans les écoles francophones de Montréal.

1.1 Les expressions figées dans l'acquisition des langues et dans la communication

La recherche sur le langage figé a connu un essor considérable grâce au développement de l'informatique. En effet, celui-ci a permis aux chercheurs d'analyser des grands corpus linguistiques et par le fait même, d'étudier la récurrence de certaines combinaisons linguistiques dans la langue. Parmi les figures éminentes de la recherche sur corpus, mentionnons Sinclair (1991), de Pawley et Syder (1983), et plus récemment, Hoey (2005), dont les travaux ont démontré la présence importance de formes figées dans la langue (en particulier les collocations). Cela dit, les EF constituent un phénomène linguistique intéressant non seulement en

raison de leur nombre et de leur variété, mais aussi à cause de leur rôle essentiel dans la communication et dans l'apprentissage des langues première et seconde. L'une des découvertes importantes concernant l'acquisition des expressions est que les expressions figées seraient apprises et emmagasinées dans la mémoire de façon holistique (globale), c'est-à-dire sans analyse grammaticale et lexicale (Wood 2002; Wray et Perkins 2000; Wray 2002, Schmitt *et al.* 2004). Au moment de l'énonciation, les locuteurs puiseraient donc les expressions figées directement dans leur mémoire, sans devoir construire un énoncé à partir de nouveaux éléments. L'emploi de EF constituerait donc une sorte de « raccourci » cognitif qui réduirait de beaucoup l'effort de traitement, comparativement à la formulation d'un énoncé non préfabriqué. Le traitement holistique des expressions figées, qui a été observé chez les locuteurs natifs et chez les locuteurs non natifs (Jiang et Nekrasova 2007), faciliterait la communication tant chez les LN que chez les LNN. Notamment, chez les apprenants L1 et L2 de niveau préscolaire, les expressions figées permettraient aux jeunes enfants d'établir une interaction verbale avec leurs éducateurs et avec les autres enfants, et ce, avant même qu'ils soient capables de produire eux-mêmes des énoncés complexes à partir de règles grammaticales (Wray 2002). En effet, les très jeunes enfants dépendent des adultes pour subvenir à leurs besoins physiques et affectifs. Ils comptent donc sur des pseudo-phrases pour communiquer, du type « maman partie », « papa dodo ». Cela dit, il a également été postulé que l'emploi d'expressions figées favoriserait l'appropriation de règles syntaxiques chez les enfants apprenant leur L1, ainsi que chez les apprenants L2 (Wong-Fillmore 1976, Myles *et al.* 1998). Autrement dit, au cours de leur acquisition, les apprenants se serviraient des EF comme modèles pour déduire des règles grammaticales et construire leurs propres énoncés. Cette hypothèse demeure cependant contestée : Krashen et Scarcella (1978) ne reconnaissent qu'un rôle minimal à l'utilisation de phrases-clés dans l'acquisition. Selon ces chercheurs, l'emploi d'expressions permettrait d'établir une communication minimale dans la langue cible chez les locuteurs non natifs de niveau débutant (exemples : « Je m'appelle X », « Comment ça va ? »), mais il ne jouerait pas de rôle significatif dans le développement de la grammaire.

Dans un autre ordre d'idées, une bonne connaissance des EF favoriserait la fluidité verbale (« fluency ») chez les locuteurs natifs et non natifs (Wood 2006; Pawley & Syder 2000) « it is the store of memorized constructions and expressions, more than anything, that is the key to nativelike fluency » (Pawley et Syder 2000, p. 195). En effet, puisqu'elles sont stockées comme une seule unité lexicale, elles sollicitent beaucoup moins d'effort de traitement au moment de la production, ce qui augmente la vitesse d'élocution, diminue la longueur et la fréquence des pauses, et permet de plus longues séquences de discours ininterrompues entre les pauses (Wood 2006). Également, il semble que l'utilisation d'EF aiderait le locuteur à structurer son discours : certains énoncés étant préfabriqués, l'apprenant peut accorder plus d'attention à la planification des autres énoncés (Raupach 1984).

Il est à noter que l'emploi d'EF facilite la communication non seulement pour l'émetteur d'un message, mais aussi pour son interlocuteur. En effet, les tournures familières sont davantage susceptibles de mener à une compréhension et à une interprétation justes de la part de celui qui reçoit le message : « just as formulaic material is easier for the speaker to encode, so also, if the hearer has a lexical entry for a wording, s/he will find it easier to decode, compared with something more novel » (Wray 2007, p.40). Par exemple, il est plus facile de comprendre la formule préfabriquée « Quelle heure est-il ? », qu'un énoncé construit à neuf, comme « À quel moment de la journée sommes-nous ? ». La première tournure est plus fréquente et plus naturelle chez les locuteurs natifs : elle est donc beaucoup plus facile à décoder que la seconde, qui bien que grammaticale, n'est pas idiomatique. D'ailleurs, il semble qu'une utilisation appropriée des EF permettrait aux apprenants de s'exprimer de façon plus idiomatique (plus « naturelle ») dans leur langue seconde : « the formulaic sequences used by native speakers are not easy for learners to identify and master, and (...) their absence greatly contributes to learners not sounding idiomatic » (Wray 2002, p. 176). La maîtrise des expressions figées serait donc essentielle pour atteindre les compétences linguistiques d'un locuteur natif.

En somme, la maîtrise des expressions figées constitue un atout pour tous les locuteurs du français, peu importe leur âge et leur niveau de compétences

langagières. En plus de faciliter le traitement cognitif du langage, elles constituent une composante essentielle de la communication, de l'acquisition et de l'idiomaticité d'une langue. C'est pourquoi il est important de s'interroger sur les facteurs qui sont susceptibles de favoriser, ou au contraire, de défavoriser, le développement des connaissances des enfants en matière d'expressions figées.

1.2 Statut particulier des élèves allophones de Montréal

Pour éviter toute confusion, il nous paraît convenable de définir d'emblée ce que nous entendons par « allophone » dans le cadre de cette étude. En effet, on confère à certaines personnes le statut d'allophone en fonction de la ou des langues officielles du territoire concerné. L'Office québécois de la langue française (OQLF) définit un allophone comme étant une « personne dont la langue d'usage (celle qui est parlée à la maison) ou la langue maternelle est différente de la ou des langues officielles du pays où elle se trouve »¹. Au Canada, les deux langues officielles sont l'anglais et le français. Donc, à l'échelle canadienne, les personnes dont la langue d'usage n'est ni l'anglais ni le français sont considérées comme des allophones.

Or, les deux langues officielles du Canada n'ont pas le statut de langue officielle dans toutes les provinces du pays. Au Québec, seul le français a été adopté comme langue officielle. Comme cette recherche a été effectuée exclusivement en territoire québécois et, de surcroît, dans une école francophone, les élèves ayant déclaré une langue d'usage autre que le français ont été considérés comme des allophones dans notre étude.

Par ailleurs, soulignons que dans la définition proposée par l'OQLF, la langue d'usage d'une personne équivaut à sa langue première (ou maternelle). Dans la présente étude, la langue d'usage (parlée à la maison) a également été considérée comme la langue première (ou maternelle) des participants. Ce choix a été fait en raison de la complexité des contextes linguistiques dans les milieux familiaux des

¹ Définition du Grand dictionnaire terminologique, dictionnaire en ligne relevant de l'Office québécois de la langue française. Adresse Internet : <http://www.granddictionnaire.com>. Consulté le 11/02/2008.

élèves allophones. En effet, il peut être difficile de déterminer quelle est la langue dominante d'un enfant, notamment dans le cas où les parents ne partagent pas la même langue première, ou dans le cas des familles où se côtoient deux ou plusieurs langues d'usage. Par conséquent, il s'est avéré nécessaire de suivre une ligne directrice pour déterminer quels enfants seraient considérés comme des allophones dans cette recherche. Les critères d'identification des élèves allophones seront expliqués plus en détail dans la section *Méthodologie*.

1.3 Élèves allophones à Montréal : survol d'une démographie scolaire

Chaque année, le Québec accueille de nombreux immigrants venus de partout dans le monde. Bien que la plupart d'entre eux (61 %) parlent français à leur arrivée au Québec², une proportion considérable des nouveaux arrivants ne parle pas cette langue. La majorité des immigrants, soit environ 78,7 %, s'établissent dans la région Montréal³. C'est donc dans la population métropolitaine que l'on observe la plus grande concentration d'allophones.

La pluralité linguistique se reflète dans tous les secteurs, et les établissements scolaires n'y font pas exception. Bien que le français demeure la langue première la plus répandue dans les établissements francophones de Montréal, les enfants allophones forment désormais une proportion très importante de la clientèle scolaire. En 2006-2007, le ministère de l'Éducation, du Loisir et du Sport (MELS) a recensé, au total, 179 000 élèves répartis dans les écoles primaires et secondaires

² Source : Immigration et communautés culturelles. Statistiques du 3^e trimestre de 2007. Site internet : <http://www.micc.gouv.qc.ca/publications/fr/recherches-statistiques/BulletinStatistique-2007trimestre3-ImmigrationQuebec.pdf>. Consulté le 05/02/2008. Nota: La région métropolitaine de Montréal inclut les agglomérations de Montréal, Laval et Longueuil.

³ Source : Immigration et communautés culturelles. Statistiques du 3^e trimestre de 2007. Site internet : <http://www.micc.gouv.qc.ca/publications/fr/recherches-statistiques/BulletinStatistique-2007trimestre3-ImmigrationQuebec.pdf>. Consulté le 05/02/2008. Nota: La région métropolitaine de Montréal inclut les agglomérations de Montréal, Laval et Longueuil.

francophones de la région de Montréal. Sur ce nombre, 99 000 élèves (55,4 %) avaient le français comme langue première. Parmi les autres élèves, 11 000 avaient l'anglais comme langue première (6 %), contre 69 000 élèves (38,5 %) dont la langue première n'était ni le français ni l'anglais⁴. Devant ces statistiques, force est de constater que près de la moitié (44,5 %) des élèves qui fréquentent le système scolaire francophone de Montréal ont une langue première autre que le français.

La commission scolaire de Montréal se démarque non seulement par sa forte proportion d'allophones, mais aussi par son impressionnante diversité linguistique. Pour l'année scolaire 2005-2006, la commission scolaire de Montréal (CSDM) a recensé pas moins de 148 langues maternelles et 188 lieux de naissance, et ces chiffres augmentent d'une année à l'autre⁵. Le MELS donne également accès à des statistiques intéressantes en ce qui concerne les divers groupes linguistiques des écoles francophones de Montréal. En 2006-2007, les arabophones et les hispanophones étaient les deux premiers groupes linguistiques en importance après les francophones, avec 15 000 et 12 000 élèves respectivement, surpassant ainsi le nombre d'anglophones, qui s'élevait quant à lui à près de 11 000. Parmi les autres langues maternelles qui sont fortement représentées dans le système scolaire français, on compte le créole (5600 élèves), le chinois (5000 élèves), le vietnamien (3000 élèves) et le russe (2200 élèves). À l'image des autres écoles montréalaises, l'école où s'est déroulée la collecte de données présente une très forte concentration d'élèves allophones, issus de différentes communautés linguistiques et ethniques. Le créole et le mandarin sont les langues premières les plus fréquentes parmi les élèves de l'école⁶. Cela dit, signalons que les allophones qui fréquentent une classe régulière dans une école francophone sont en situation de *submersion* : ils sont

⁴ Source : Statistiques du ministère de l'Éducation, du Loisir et du Sport. Site internet : http://www.mels.gouv.qc.ca/stat/stat_det/PPS_eff.htm. Consulté le 11/02/2008.

⁵ Source : Commission scolaire de Montréal. Site internet : <http://www.csdm.qc.ca/Csdm/Administration/default.aEF?csdm=mosaique>. Consulté le 07/02/2008.

⁶ Communication écrite avec la conseillère pédagogique de l'école, le 13/02/2008.

scolarisés dans une langue autre que leur langue première. Le contexte de *submersion* existe par opposition au contexte d'*immersion*, où la langue apprise a un statut de langue seconde, et qui est par conséquent enseignée comme une langue seconde.

La population allophone à Montréal connaît une croissance constante et forme désormais une proportion considérable de la métropole. Étant donné cette situation, il devient important de s'intéresser à l'éducation des jeunes allophones, et plus particulièrement en ce qui concerne leurs aptitudes en français. Effectivement, leur niveau de compétence dans la langue de scolarisation pourrait s'avérer décisif pour leur succès non seulement dans la poursuite de leurs études, mais aussi pour leur intégration dans une société majoritairement francophone. Ajoutons également que des citoyens allophones compétents en français constituent un atout pour la société québécoise, soucieuse de préserver son caractère francophone.

1.4 Développement des connaissances linguistiques chez les allophones et les francophones

La maîtrise d'une langue nécessite un ensemble de connaissances et de compétences, dont la lecture, l'écriture et la grammaire, mais aussi les connaissances lexicales. Alors que la lecture et l'écriture sont généralement acquises à l'école, l'apprentissage du lexique débute alors que l'enfant n'est âgé que de quelques mois, et évolue très rapidement pendant les premières années de sa vie, période pendant laquelle l'évolution de ses habiletés langagières dépend surtout de l'interaction avec les parents. À l'âge de 6 ans, les enfants ont un vocabulaire d'environ 2000 à 2500 mots (Blaye et Lemaire 2007). Lorsqu'ils arrivent à l'école, les enfants francophones possèdent donc, en principe, des connaissances linguistiques suffisantes pour entamer leur scolarisation. La situation est cependant différente pour les enfants issus d'un milieu familial où la langue d'usage n'est pas le français. Ayant été peu ou pas du tout exposés au français pendant la période préscolaire, ils entrent à l'école avec des connaissances lexicales plus restreintes que celles de leurs camarades francophones, qui ont été exposés au français depuis leur naissance. Ces lacunes peuvent entraîner des répercussions sur l'acquisition des

compétences en littératie. En effet, un vocabulaire pauvre ralentit la lecture, nuit à la compréhension et limite la capacité d'apprendre de nouveaux mots. À l'inverse, un vocabulaire diversifié assure aux enfants non seulement une rapidité accrue et une meilleure compréhension en lecture, mais aussi l'acquisition de nouveaux mots (par inférence). D'ailleurs, des études réalisées auprès d'élèves allophones et francophones confirment que les enfants allophones ou nés de parents allophones avaient des résultats plus faibles en écriture et en lecture que les enfants de parents francophones.

Notamment, une étude menée par le ministère des Ressources humaines et du Développement social du Canada visait à évaluer les compétences en lecture d'élèves âgés de 15 ans (test PISA 2000). Les élèves étaient classés en trois groupes linguistiques : les élèves de souche (francophones nés au Québec), les élèves de première génération (nés au Québec, mais dont les parents étaient nés à l'étranger) et les élèves immigrants (nés à l'étranger). La langue parlée à domicile s'est révélée un facteur important dans la performance des élèves : les élèves qui parlaient français à domicile (peu importe leur statut linguistique) ont mieux réussi que les élèves qui parlaient une autre langue à la maison.

Dans la même veine, une étude longitudinale réalisée par Statistique Canada (2001) traite du rendement scolaire des enfants en fonction de la langue maternelle des parents. Les participants à l'étude étaient divisés en trois groupes linguistiques de la façon suivante : les enfants de parents nés au Canada, les enfants de parents immigrants dont la langue maternelle est l'anglais ou le français, et les enfants de parents immigrants dont la langue maternelle n'est ni l'anglais ni le français. Les élèves ont été évalués à l'aide d'une entrevue avec les parents, une entrevue avec l'enseignant ainsi que des tests formels. Les élèves dont les parents n'avaient ni le français ni l'anglais comme langue maternelle ont eu des résultats beaucoup plus faibles que leurs camarades francophones en lecture et en écriture, surtout au début de leur parcours scolaire. Toutefois, la plupart ont fait des progrès d'année en année et ont rattrapé les autres élèves vers l'âge de 13 ans.

Donc, il semble que les élèves allophones qui ne parlent pas français à la maison soient désavantagés par rapport à leurs pairs francophones au plan des

compétences en français dans les premières années de leur scolarisation ; or, cet écart aurait tendance à s'estomper au fil des années. Cela dit, même si les allophones rattrapent généralement leurs camarades francophones en lecture et en écriture après un certain nombre d'années passées dans le système scolaire francophone, il est possible que leurs compétences en matière d'expressions figées demeurent inférieures à celles des francophones au fil du temps. En effet, les expressions figées font partie des subtilités de la langue qui permettent de distinguer entre les locuteurs natifs et les locuteurs non natifs avancés : « *skilful use of formulaic sequences is generally considered to come late in the acquisition process, and it is not unreasonable to question whether mastery can ever truly be achieved. Indeed, control of formulaic sequences is one element than can normally be relied upon to distinguish between natives and even relatively advanced nonnatives* » (Conklin & Schmitt 2007, p. 13). Donc, il se peut que les enfants francophones se distinguent de leurs pairs allophones en ce qui a trait à la connaissance des expressions figées, et ce, même si les enfants allophones se débrouillent bien dans d'autres disciplines du français.

1.5 Conclusion du chapitre

Ce chapitre visait à situer notre recherche dans son contexte sociohistorique et d'en démontrer la pertinence dans un contexte montréalais plurilingue. Dans cette optique, nous avons exposé la situation linguistique des élèves allophones dans leur milieu scolaire. L'accent a été mis sur l'importance des connaissances linguistiques, qui s'avèrent décisives pour la réussite scolaire de tous les élèves et pour l'intégration sociale des allophones. Le contexte de la recherche ayant été cerné, nous aborderons, dans le prochain chapitre, la partie théorique de notre recherche.

CHAPITRE II

CADRE THÉORIQUE

Ce chapitre théorique se divise en trois parties. Dans la première partie, nous ferons un survol terminologique : cette mise au point est importante, car la terminologie se rapportant à la phraséologie est abondante, et les nuances qui existent entre les termes ne sont pas toujours clairement établies. En effet, les définitions d'un même terme diffèrent grandement d'un ouvrage à l'autre, et parfois même à l'intérieur d'une même source. Nous proposons donc au lecteur un survol des termes les plus fréquemment employés pour désigner les suites figées en français, à savoir les collocations, les séquences préfabriquées, les expressions idiomatiques, les locutions et les expressions figées¹ Pour chaque terme, nous donnerons la ou les définitions qui nous semblaient les plus appropriées, et nous donnerons quelques exemples pour illustrer chacune d'elles. Au terme de la section terminologique, nous présenterons notre définition des expressions figées telles que nous les concevons dans le cadre de notre étude. Dans la deuxième partie de ce chapitre, nous poursuivrons notre description des expressions figées avec la présentation des caractéristiques servant à l'identification des formes figées dans la langue. Nous verrons que certaines de ces caractéristiques, dont la fréquence et le

¹ Pour une liste plus complète de la terminologie employée par les chercheurs, nous invitons le lecteur à se référer aux annexes.

degré d'opacité sémantique, ont un effet important sur l'acquisition des expressions figées. Quant à la troisième partie du chapitre, elle sera consacrée à l'exploration des enjeux liés à l'apprentissage et à l'utilisation du langage figé dans l'acquisition d'une langue première et d'une langue seconde. La discussion portera essentiellement sur le rôle des expressions figées à chaque étape de l'acquisition d'une langue première et seconde et sur l'influence de différents facteurs individuels sur la capacité de comprendre et de produire des expressions figées.

2.1 Tour d'horizon du domaine de la phraséologie : quelques repères terminologiques pour s'y retrouver

2.1.1 Collocations

De façon générale, le terme *collocation* se réfère à une suite lexicale, formée par une association syntaxique et sémantique plus ou moins forte entre deux mots ou plus. Les collocations concernent donc la combinatoire des mots, c'est-à-dire la façon dont les mots de différentes catégories grammaticales s'agencent dans une langue donnée. Le GDT définit la collocation de la façon suivante : « présence d'au moins deux unités linguistiques distinctes dans un énoncé, qui sont liées par un rapport de proximité syntaxique et de relative dépendance ». Hoey (2005), pour sa part, propose une définition qui met en relief le rôle essentiel de la fréquence dans la formation des collocations : « it is a psychological association between words (rather than lemmas) up to four words apart and is evidenced by their occurrence together in corpora more often than is explicable in terms of random distribution ». (p.5). Pour illustrer ces définitions, examinons les collocations formées à partir du mot *eau*, telles que recensées dans le *Dictionnaire des cooccurrences* de Beauchesne (2001), qui propose environ 90 adjectifs et une quarantaine de verbes typiquement associés à l'eau (p. 120). Parmi les adjectifs, citons *calme*, *cristalline*, *diaphane*, *fraîche*, *intarissable*, *jaillissante*, *ondoyante*, *pluviale*, *potable*, *saumâtre*, *translucide*, *trouble* et *vive*. Quant aux verbes, mentionnons *bouillonner*, *cascader*, *clapoter*, *jaillir*, *onduler*, *ruisseler*, *se troubler* et *tourbillonner*. Les combinaisons suivantes forment donc des collocations : *eau cristalline*, *eau fraîche*, *eau trouble* ; *l'eau jaillit*, *l'eau*

cascade, l'eau ruisselle, etc. Dans le cas des collocations formées à partir du nom eau (*eau* + adjectif / *eau* + verbe), le degré de figement est relativement faible, puisqu'on peut associer le mot *eau* à une multitude d'adjectifs et de verbes, et à l'inverse, les adjectifs et les verbes qui sont typiquement associés à l'eau peuvent se retrouver en compagnie d'autres noms. Toutefois, le fait que le mot « eau » se retrouve fréquemment en compagnie des adjectifs et des verbes précités témoigne d'une réciprocité entre ces mots : il y a donc un figement, si faible est-il. Cela dit, il existe des collocations où le figement est plus marqué : c'est le cas par exemple de l'adjectif « aquilin », qui est fortement associé au mot « nez », ou du verbe « hocher », qui est étroitement lié au nom « tête ». Le terme *collocation* est un terme général qui couvre un très grand nombre de suites lexicales. Ce phénomène linguistique est donc propice aux grandes analyses de corpus, où l'utilisation d'un ordinateur est cruciale pour identifier les mots qui se retrouvent ensemble de façon récurrente (voir par exemple les recherches de Sinclair 1991 et Hoey 2005). Les collocations ont déjà fait l'objet d'études acquisitionnelles chez des adultes en langue seconde, et il a été prouvé qu'elles posaient des difficultés même aux apprenants les plus avancés (voir par exemple Nesselhauf 2005). Il serait donc extrêmement difficile d'évaluer la connaissance des collocations dans une population d'enfants, surtout si celle-ci comprend une forte proportion d'allophones, comme c'est le cas dans cette recherche. C'est pourquoi notre évaluation portera plutôt sur un autre type de suites lexicales, les expressions figées, (qui seront définies ultérieurement), celles-ci étant davantage ancrées dans l'usage et présentant des associations lexicales très fortes.

2.1.2 Séquence préfabriquée

Le terme « séquence préfabriquée »² est un terme générique qui fait référence à l'ensemble des phénomènes relatifs à la phraséologie. Erman et Warren (2000) en donnent la définition suivante : « A prefab is a combination of at least two words favoured by native speakers in preference to an alternative combination which could

² Terme forgé par (Forsberg 2006). Termes anglais équivalents: *formulaic sequence* (Wray 2002), *prefab* (Erman et Warren 2000).

have been equivalent had there been no conventionalisation » (Erman et Warren 2000, p.31). En effet, bien qu'il existe plus d'une façon d'exprimer un concept dans une langue donnée, les locuteurs natifs optent naturellement pour les formulations conventionnelles. Par exemple, lors d'une conversation téléphonique, un locuteur natif sera porté à employer des formules standards, comme « oui, allô ? » et « un instant, s'il vous plaît », plutôt que des formules non courantes comme « j'écoute » ou « attendez un peu ». Cela dit, nous présenterons aussi une autre définition des séquences préfabriquées, celle de Wray (2002). Celle-ci a été élaborée dans une perspective psycholinguistique et met l'accent sur le traitement cognitif des séquences préfabriquées : « a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar » (Wray 2002, p. 9). Donc, les formules comme « oui, allô » et « un instant, s'il vous plaît » sont mémorisées par le locuteur et produites comme des automatismes : elles ne sont pas créées au moment de l'énonciation. Les processus cognitifs mis en œuvre lors du traitement des expressions seront explorés ultérieurement. Du côté de la recherche acquisitionnelle française, citons la définition de Forsberg (2006). Elle est essentiellement fondée sur celle d'Erman et Warren (2000) et de Wray (2002), mais elle introduit une distinction importante entre les séquences préfabriquées qu'emploient les LN et celles qu'emploient les LNN :

une séquence préfabriquée est composée d'au moins deux mots graphiques. Elle est préférée, c'est-à-dire plus fréquente dans un contexte donné, chez les locuteurs natifs, qu'une (sic) autre combinaison qui aurait pu être équivalente si la première n'avait pas été conventionnalisée. Or le répertoire des séquences préfabriquées chez les locuteurs L2 diverge de celui des locuteurs natifs. Par conséquent, une séquence préfabriquée peut aussi être une séquence préférée par un locuteur L2, identifiée à cause de son emploi idiosyncrasique (morphosyntaxique, sémantique, pragmatique) et récurrent dans l'interlangue d'un locuteur L2. (Forsberg 2006, p. 44).

Donc, les locuteurs natifs adultes maîtrisent et utilisent un certain nombre de séquences préfabriquées conventionnelles, par exemple : *Comment ça va ? Il fait*

chaud. On verra bien. Je suis désolé(e). (Exemples de Forsberg 2006). En ce qui a trait aux locuteurs non natifs, ils emploient également des séquences préfabriquées, mais elles ne sont pas nécessairement les mêmes que celles employées par les LN. Elles peuvent être employées dans un sens différent de leur sens conventionnel (par exemple : « C'est chaud.* », employée dans le sens de « J'ai chaud. »), ou être utilisées de façon « trop fixe », par exemple : « Monique j'adore le tennis.* » (au sens de « Monique adore le tennis » dans l'étude Myles, Hooper et Mitchell 1998).

2.1.3 Expressions idiomatiques

Les expressions idiomatiques³ se distinguent des autres expressions par leur sens métaphorique. Il existe un très grand nombre d'expressions idiomatiques dans la majorité des langues ; le français en compterait plus de 2300 (Denhière et Verstigel 1997). Examinons une première description de ce type d'expression :

Ces locutions ne sont ni des proverbes, ni des dictons, bien qu'elles soient souvent qualifiées, comme ces derniers, de familières ou populaires, encore que les plus grands écrivains les utilisent, sans pour autant être à court d'inspiration, car elles sont porteuses de sens difficile, voire impossible à transmettre par de simples paraphrases littérales. Beaucoup d'entre elles en effet, mais pas toutes, présentent un caractère métaphorique, voire sont parfois quasiment des métaphores. [...] Mais pour être considérée comme idiomatique, une locution doit être reconnaissable, identifiable comme telle, c'est-à-dire qu'elle doit posséder un certain degré de figement, ce qui la différencie d'une métaphore proprement dite. S'opposent en effet la productivité ou la créativité métaphorique au gel ou au figement idiomatique : alors qu'on peut toujours créer de nouvelles métaphores, les expressions idiomatiques constituent un corpus que l'on peut consigner dans les dictionnaires. (Denhière et Verstigel 1997, p. 119).

Le dictionnaire anglais *Cobuild Dictionary of Idioms* (1995, p. iv) offre également une définition éclairante : «An idiom is a special kind of phrase. It is a group of words which have a different meaning when used together from the one it would have if the meaning of each word were taken individually. [...] Idioms are typically metaphorical: they are effectively metaphors which have become 'fixed' or 'fossilized'» (Svensson

³ Terme anglais : *idiom* ou *idiomatic expression*.

2004; p. 21). Cette définition fait référence non seulement au caractère métaphorique des expressions idiomatiques, mais aussi à leur non compositionnalité, c'est-à-dire qu'on ne peut pas déduire le sens de ces expressions à partir du sens des mots qui les composent (ce concept sera expliqué plus en profondeur dans la prochaine section). L'aspect métaphorique et la non compositionnalité des expressions idiomatiques sont les caractéristiques les plus souvent citées par les chercheurs pour décrire ce type d'expression. Voici quelques exemples d'expressions idiomatiques, extraites du Dictionnaire des expressions et locutions de Rey et Chantreau (2003) : *mettre les bâtons dans les roues à quelqu'un*, au sens de « susciter des difficultés, chercher à gêner une entreprise ». *Prendre le taureau par les cornes*, au sens de « s'attaquer à la difficulté avec détermination ». *Vivre d'amour et d'eau fraîche*, au sens de « se contenter des sentiments et ne pas se préoccuper des nécessités matérielles ».

Étant donné leur nature métaphorique, elles appellent à une interprétation figurée ou imagée. Alors que les locuteurs natifs d'âge adulte comprennent et utilisent un grand nombre d'expressions idiomatiques, il s'avère généralement difficile pour les jeunes enfants de saisir le sens de ces expressions. En effet, la capacité des enfants d'interpréter le sens figuré n'est pas entièrement développée, ce qui limite leur compréhension des expressions idiomatiques. Nous expliquerons plus loin en quoi le sens figuré influence l'acquisition des expressions et quels sont les autres facteurs qui interviennent dans l'interprétation du sens.

Malgré que le sens métaphorique et la non compositionnalité soient typiquement associés aux expressions idiomatiques, il n'en reste pas moins que d'autres types d'expressions, qui sont généralement plus littérales, (par exemple, les expressions figées et les locutions, que nous aborderons plus loin), puissent aussi présenter un figement sémantique ainsi qu'un aspect métaphorique, quoique dans un degré moindre. Par exemple, les expressions « à vol d'oiseau » et « perdre le fil » ont un sens figuré, mais la métaphore est moins forte que dans des expressions idiomatiques proprement dites.

2.1.4 Locutions

Les locutions⁴ sont généralement décrites comme des syntagmes occupant une fonction syntaxique précise au niveau de la phrase, à même titre que les mots simples : elles peuvent tenir lieu de noms, d'adjectifs, de verbes, d'adverbes, de prépositions et de conjonctions. Présentons d'abord la définition élaborée par Gross (1996) : « une locution est un syntagme (nominal, verbal, adjectival, adverbial) dont les éléments composants ne sont pas actualisés individuellement et qui forme un concept autonome, que le sens global soit figé ou non. On parlera aussi de catégorie complexe ou polylexicale ». (p. 154). Donc, selon la définition de Gross, les locutions sont des expressions qui sont figées grammaticalement, mais pas nécessairement sémantiquement. Une autre définition est proposée dans le *Dictionnaire d'expressions et locutions* de Rey et Chantreau (2003) : « une unité fonctionnelle plus longue que le mot graphique, appartenant au code la langue (devant être apprise) en tant que forme stable et soumise aux règles syntactiques de manière à assumer la fonction d'intégrant (au sens de Benveniste) » (p. IX). À la lumière de ces définitions, on peut constater que les locutions constituent un phénomène essentiellement grammatical, où l'aspect sémantique est relégué au second plan. Voici quelques exemples de locutions : tant bien que mal (locution adverbiale), remettre en question (locution verbale), à proximité de (locution prépositive). Fait intéressant à noter : dans certaines catégories syntaxiques du français, les formes figées sont plus fréquentes que les mots simples. Selon les travaux du LADL⁵, on compte 300 000 à 400 000 locutions nominales contre 80 000 noms simples, 20 000 locutions verbales contre 8000 à 12 000 verbes simples, et 6000 locutions adverbiales contre 2000 adverbes simples (Chanier 1997).

⁴ Terme anglais : *phrase*.

⁵ Laboratoire d'analyse documentaire et linguistique (Université Paris 7).

2.1.5 Expression figée

Les expressions figées, comme leur nom l'indique, sont caractérisées par un figement. Ce figement peut porter sur le sens de l'expression, sur la syntaxe ou sur le lexique. Gross (1996) définit les expressions figées de la façon suivante :

Une séquence est figée du point de vue syntaxique quand elle refuse toutes les possibilités combinatoires ou transformationnelles qui caractérisent habituellement une suite de ce type. Elle est figée sémantiquement quand le sens est opaque ou non compositionnel, c'est-à-dire quand il ne peut pas être déduit du sens des éléments composants. Le figement peut être partiel si la contrainte qui pèse sur une séquence donnée n'est pas absolue, s'il existe des degrés de liberté (Gross 1996, p. 154).

Donc, les expressions sont « figées » en raison d'un ensemble de contraintes plus ou moins rigides d'ordre syntaxique ou sémantique. Les « possibilités combinatoires et transformationnelles » se réfèrent à des opérations syntaxiques qui permettent de détecter, dans un syntagme, la présence d'un figement. Quelques exemples d'opérations syntaxiques sont la passivation, la pronominalisation, la relativation, l'interrogation et le changement de déterminant.

Voici un exemple d'expression figée sémantiquement et syntaxiquement : *prendre le large* au sens de « s'enfuir » (Rey et Chantreau 2003). Il y a un figement sémantique parce que le sens global, c'est-à-dire « s'enfuir », ne peut être interprété à partir des mots composant cette expression. Cette expression est figée syntaxiquement, car elle ne peut être soumise aux transformations syntaxiques suivantes :

Passivation : * le large a été pris

Pronominalisation : *elle l'a pris

Relativisation : *le large qu'elle a pris

Interrogation : Qu'a-t-elle pris ?

Changement de déterminant : * elle a pris (*un, *des, *ce, *plusieurs) large (s).

Voici d'autres exemples d'expressions figées : *jour pour jour*, *prêter main-forte*, *la huitième merveille du monde*, *avoir deux poids et deux mesures*, *noir sur blanc*, *bon premier*.

Il est important de signaler que les frontières entre ces types d'expressions ne sont pas toujours nettes ; les catégories peuvent parfois se recouper. Par exemple, *vendre la mèche* pourrait être perçue à la fois comme une expression idiomatique (puisque'elle a un sens figuré), une locution verbale (puisque'elle fonctionne comme un verbe) ou une collocation (dans la mesure où le verbe « vendre » est également associé à d'autres compléments directs). Cela dit, ce que l'on doit retenir de cette première section, c'est que toutes les expressions, de façon générale, forment soit une unité sémantique (ce qui est particulièrement évident dans le cas des expressions idiomatiques) soit une unité syntaxique (ce qui est plus saillant dans le cas des locutions), ou les deux (ce qui est souvent le cas des expressions figées). Mentionnons également que certaines expressions sont plus fixes dans leur forme (expressions idiomatiques), tandis que d'autres sont plus flexibles (collocations). Enfin, soulignons que toute expression, peu importe son degré de figement et ses caractéristiques sémantiques ou syntaxiques, doit nécessairement être récurrente dans la langue parlée ou dans la langue écrite. Dans la prochaine section, les notions de fréquence, de figement, ainsi que les particularités sémantiques, syntaxiques et lexicales seront analysées plus en profondeur. Nous explorerons également d'autres caractéristiques, qui n'ont pas été évoquées jusqu'à présent, mais qui sont également importantes pour décrire les expressions figées et pour expliquer les processus menant à leur acquisition.

2.1.6 Vers une définition des expressions figées

Dans le cadre de cette recherche, c'est le terme « expression figée » qui a été adopté. Bien que la définition proposée par Gross (1996) mette en lumière quelques-unes des caractéristiques essentielles de ce type d'expression, nous proposons une nouvelle définition fondée en grande partie sur celle de Gross (1996), mais à laquelle nous apportons quelques précisions complémentaires afin de mieux cerner le phénomène qui nous intéresse : *une expression figée est une suite d'au moins deux mots graphiques et occupe, à même titre que les mots simples, une fonction grammaticale précise au niveau de la phrase. Elle est caractérisée par un figement — partiel ou total — au plan syntaxique, lexical ou sémantique. Elle fait partie*

intégrant le vocabulaire d'usage des locuteurs natifs d'une langue. Cette définition a l'avantage de tenir compte des particularités orthographiques et grammaticales, des différentes formes possibles de figement, ainsi que de la place prépondérante des expressions figées dans la langue d'usage. Toutefois, nous sommes consciente que cette définition n'offre – comme toute autre définition des expressions figées d'ailleurs – qu'une représentation assez imprécise du phénomène envisagé. C'est pourquoi il est crucial de compléter la description des expressions figées à l'aide de caractéristiques spécifiques aidant à l'identification des expressions dans la langue. En effet, l'identification des expressions est une tâche délicate. En effet, ce qui est considéré comme du langage figé par certains membres d'une communauté linguistique n'est pas nécessairement perçu comme tel par tous les membres : certaines expressions sont connues de presque tous les locuteurs natifs ou quasi natifs ; d'autres sont beaucoup moins connues (Erman et Warren 2000). De plus, certaines expressions peuvent facilement passer inaperçues : à première vue, certaines ressemblent à des combinaisons lexicales libres, alors qu'en réalité elles présentent un figement (Erman et Warren 2000).

Dans cette section, nous avons réalisé une synthèse des principales caractéristiques (ou critères d'identification) des expressions que l'on retrouve dans les recherches en phraséologie. Il s'avère que peu importe le type d'expression étudié (expressions idiomatiques, expressions figées, locutions, etc.), les caractéristiques évoquées sont sensiblement les mêmes. Rappelons que les nuances entre les types d'expressions ne sont pas toujours tranchées ; par conséquent, les critères d'identification s'appliquent généralement à l'ensemble des expressions. À des fins organisationnelles, nous avons classé les caractéristiques dans cinq catégories : caractéristiques cognitives, caractéristiques liées à l'usage, caractéristiques syntaxiques, caractéristiques lexicales, caractéristiques sémantiques et caractéristiques phonologiques.

2.2 Caractéristiques des expressions figées

2.2.1 Caractéristiques cognitives

Parmi les caractéristiques souvent mentionnées dans littérature sur le langage figé, citons la **mémorisation** (Wray 2002; Svensson 2004). En effet, les expressions figées sont apprises, mémorisées et employées de façon spontanée par les locuteurs. Elles ne sont donc pas « générées » par les locuteurs en temps réel au moment de l'élocution. Rappelons à cet effet la définition de *formulaic sequence*, telle que proposée par Wray (2002) : « a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar » (Wray 2002, p. 9). Cette définition fait également référence au **stockage holistique** des expressions dans le lexique mental. En effet, il est généralement admis chez les chercheurs que les expressions sont emmagasinées comme des unités linguistiques dans le lexique mental des locuteurs (voir par ex. Wood 2002; Wray et Perkins 2000; Wray 2002; Schmitt 2004). Bien qu'il ne soit pas possible de vérifier cette hypothèse directement, des expériences menées auprès de LN et de LNN tendent à la confirmer. Notamment, Underwood *et al.* (2004) ont pu comparer le temps de traitement d'expressions figées par rapport au temps de traitement de syntagmes non figés en étudiant le mouvement des yeux d'adultes LN et LNN. Les résultats ont permis de conclure que chez les LN, le temps de traitement était plus court lors de la lecture d'une séquence préfabriquée que d'un syntagme non figé. Les chercheurs Jiang et Nekrasova (2007) ont quant à eux soumis des LN et des LNN à des jugements de grammaticalité portant sur des expressions figées et des syntagmes non figés, et ont comparé le temps de réaction pour ces deux types de séquences. Il s'est avéré que le temps nécessaire au jugement de grammaticalité était plus rapide lorsqu'il s'agissait d'expressions figées que lorsqu'il s'agissait de syntagmes non figés, et ce, tant chez les LN que chez les LNN. Le fait que le temps nécessaire au traitement cognitif d'une expression figée soit inférieur à celui que requiert un syntagme non figé porte à croire que les expressions figées sont effectivement mémorisées et traitées de

façon holistique. Bien que la présente recherche ne soit pas axée sur les processus cognitifs menant à l'acquisition des EF, nous croyons que les expressions sont mémorisées et stockées de manière holistique, et cela aura un impact sur le type d'instrument de mesure utilisé (voir *Méthodologie*).

2.2.2 Caractéristiques liées à l'usage

La **fréquence** est très souvent considérée comme un critère d'identification des expressions dans la langue (caractéristique évoquée dans nombre de recherches, dont Wray 2002, 2007; Schmitt *et al.* 2004). Plus une combinaison de mots est fréquente dans la langue, plus il est probable qu'il s'agisse d'une EF. Par exemple, dans les corpus étudiés aux fins de cette recherche, l'expression « tout le monde » était extrêmement fréquente ; et s'est effectivement avérée être une expression figée après vérification dans *Le Robert des expressions et locutions*. Toutefois, comme le souligne Wray (2002), la fréquence à elle seule ne peut déterminer avec certitude si un groupe de mots donné constitue une expression figée. Par exemple, « aller au restaurant » peut se révéler très fréquent, mais il ne s'agit pourtant pas d'une expression figée dans la plupart des contextes. Inversement, l'expression « depuis belle lurette » est plutôt rare, mais il s'agit bien d'une expression figée (elle a le statut de locution dans *Le Petit Robert*). Cela dit, Wray (2007) mentionne qu'il faut porter une attention particulière aux facteurs susceptibles de biaiser la perception de la fréquence : la longueur du syntagme, le nombre d'occurrences nécessaires pour considérer l'expression comme étant « fréquente », et la flexibilité des paramètres de recherche en ce qui a trait aux expressions semi-fixes (c.-à-d. les EF permettant une variation restreinte, un choix ouvert ou une extension). Il est donc important de recourir également à d'autres indicateurs, par exemple la présence d'un blocage syntaxique, lexical ou sémantique (dont nous discuterons plus loin).

Cela dit, la fréquence est un facteur de premier ordre dans l'acquisition des formes figées : les locuteurs sont plus susceptibles de connaître les expressions fréquentes que les expressions rares. Plusieurs chercheurs se sont intéressés au rôle de la **familiarité** ou **fréquence d'exposition** sur l'acquisition des expressions, par exemple Laval (2003), Levorato et Cacciari (1992), Nippold et Rudzinski (1993),

Nippold et Taylor (1995; 2002). Pour sa part, Laval (2003) a exploré le rôle de la familiarité sur la compréhension d'expressions idiomatiques chez des enfants de 6 et 9 ans et des adultes locuteurs natifs du français. Les participants étaient soumis à une tâche de complétion d'expressions idiomatiques, suivie d'une tâche d'explication. Les résultats ont montré que la familiarité avait joué un rôle important dans la performance des enfants de 9 ans et des adultes (mais pas chez les enfants de 6 ans). Quant à Nippold et Rudzinski (1993), elles ont examiné le rôle de la familiarité dans une tâche d'explication du sens d'expressions idiomatiques chez des enfants et des adolescents de 10, 11 et 17 ans. Les participants ont eu davantage de facilité à expliquer les expressions très familières et familières que les expressions non familières. Levorato et Cacciari (1992), pour leur part, ont exploré les effets de la familiarité sur la compréhension et la production d'expressions idiomatiques chez des enfants LN de 8 et 11 ans. La familiarité n'a joué qu'un rôle mineur dans la compréhension des expressions ; par contre, elle a eu un effet important dans la capacité des enfants de produire les expressions à l'écrit. Dans la tâche de production, les élèves devaient lire cinq courtes histoires contenant des expressions idiomatiques incomplètes (un ou plusieurs mots avaient été enlevés). La moitié des expressions étaient considérées comme familières, c'est-à-dire que les enfants étaient susceptibles de les avoir rencontrées lors de conversations, à la télévision ou dans des livres ; les autres expressions étaient jugées non familières (ce classement a été fait en fonction du jugement de 152 enseignants à l'école primaire). La tâche des enfants consistait à compléter ces expressions à l'aide du ou des mots qui leur semblaient les plus appropriés. Les chercheurs ont ensuite classé les réponses des élèves selon les trois catégories suivantes : complétion idiomatique (les réponses qui correspondaient exactement à l'expression recherchée), complétion littérale (les réponses associées à des objets concrets ayant rapport avec l'histoire), et complétion figurative (les réponses qui montraient que l'enfant comprenait ce à quoi l'expression faisait référence, mais qu'il ne connaissait pas la façon conventionnelle de l'exprimer). Les chercheuses ont constaté que la familiarité avait eu un effet significatif sur la capacité des enfants de compléter correctement les expressions. L'âge a également eu une influence significative sur la performance des enfants, les

enfants de 11 ans ayant donné davantage de réponses idiomatiques que les enfants de 8 ans.

Dans notre recherche, nous nous intéressons à une population d'enfants allophones, francophones âgés de 8 à 13 ans, et à leur capacité de produire des expressions figées à l'écrit. Puisque la familiarité a un effet positif sur la compréhension et sur la capacité de production des expressions idiomatiques chez les enfants de cet âge, il est probable que ce facteur ait également une influence dans la compréhension et la production des expressions figées chez les enfants. Nous avons donc choisi de tenir compte du rôle de la familiarité dans le choix des expressions en vue de la création de l'instrument de mesure. En effet, il a fallu choisir des expressions figées qui sont susceptibles d'être fréquentes (ou du moins, récurrentes) dans l'input auquel ils sont exposés. C'est pourquoi les expressions que nous avons sélectionnées en vue de la création de notre instrument de mesure ont été extraites de livres destinés aux enfants : puisque ceux-ci contiennent généralement des mots fréquents (Lété 2004), ils sont susceptibles de contenir des expressions figées fréquentes (la sélection des expressions sera expliquée en détail dans la *Méthodologie*). Cela dit, rappelons que les participants de nos deux groupes linguistiques sont exposés aux mêmes expressions en français écrit. Par contre, ils n'ont pas tous la même exposition au français oral : les élèves allophones parlent une autre langue que le français à la maison, alors que les francophones parlent français à la maison. Puisque leur milieu familial leur offre peu (ou pas) d'input oral en français, les enfants allophones risquent d'entendre moins d'expressions figées. Nous nous attendons donc à ce que les allophones connaissent moins d'expressions figées que les francophones.

Un autre moyen d'identifier les expressions figées d'une langue consiste à consulter des **listes ou des ouvrages publiés** (Wray 2002; 2007). En effet, certaines expressions figées deviennent lexicalisées et sont répertoriées dans des ouvrages de référence (dictionnaires d'usage, dictionnaires de collocations, ou dictionnaires d'expressions et de locutions), des listes publiées ou des corpus. Comme le souligne Wray (2007), il faut toutefois s'assurer de la validité et de la pertinence de ces ouvrages avant de s'en servir comme méthode d'identification.

Enfin, il est également possible de repérer les expressions figées dans la langue des locuteurs non natifs en raison de leur **emploi idiosyncrasique** (Forsberg 2006; Wray 2007). Plusieurs indices permettent de déceler les expressions figées dans le discours des apprenants, notamment l'existence d'un décalage entre le niveau de compétences langagières du locuteur non natif et la complexité ou la longueur des constructions grammaticales qu'il emploie. Par exemple, un locuteur non natif débutant peut être capable d'employer correctement le syntagme « s'il vous plaît », et ce, sans savoir que « si » devient « s'il » devant « il », et sans avoir appris le sens et les flexions du verbe « plaire ». Les autres manifestations idiosyncrasiques comprennent le figement excessif ou l'emploi inapproprié d'une structure syntaxique (Wray 2002).

2.2.3 Caractéristiques syntaxiques

Dans beaucoup de recherches, le **nombre de mots** est une condition nécessaire à l'identification des expressions. Le minimum de mots requis pour donner à une expression le statut d'expression figée est généralement fixé à deux mots (voir par exemple Wray 2002; Erman et Warren 2000; Forsberg 2006). Cependant, Wray (2002) reconnaît que ce critère peut parfois être arbitraire, notamment dans les cas où deux expressions équivalentes occupent une fonction pragmatique similaire, l'une ne comptant qu'un mot et l'autre, deux ou plus (ex. : « merci » par rapport à « merci mille fois », « merci bien » ou « merci beaucoup »). D'autres exceptions existent dans les langues qui permettent de fusionner deux mots ou plus, comme l'anglais (ex. : *the effect in the long term* par rapport à *the long-term effect*). Enfin, un problème se pose également dans le cas des expressions ayant deux ou plusieurs graphies possibles (ex : pause café/pause-café).

Cela dit, lorsqu'il s'agit de compter le nombre de mots d'une expression, il faut nécessairement tenir compte des règles orthographiques et typographiques de la langue en question. En français, il existe de nombreux mots dits composés, dont beaucoup sont liés par un trait d'union (du moins, selon l'orthographe traditionnelle) : garde-robe, arc-en-ciel, lave-vaisselle. L'OQLF a fixé une norme typographique

voulant que les mots liés par un trait d'union soient exempts d'espacement. Selon cette règle, le mot « abat-jour » ne doit pas contenir d'espace, et devrait donc être considéré comme un seul mot. Dans une recherche où le nombre de mots constituerait un critère d'identification des expressions figées, le mot « abat-jour » ne devrait pas avoir le statut d'EF.

Un moyen de vérifier si un syntagme constitue une expression figée est de vérifier s'il y a un **blocage grammatical** (Svensson 2004), caractéristique aussi connue sous le nom de « blocage des caractéristiques transformationnelles » (Gross 1996) ou « *blocking of syntactic variability* » (Erman et Warren 2000). La présence d'un blocage grammatical peut être détectée en soumettant le syntagme à diverses opérations syntaxiques. Par exemple, dans les expressions nominales, adjectivales et adverbiales, il peut être impossible d'insérer un ou plusieurs mots (Gross 1996; Svensson 2004), d'en effacer (Svensson 2004) ou d'en inverser l'ordre sans perdre le sens global de l'expression. Par exemple, il est impossible d'intervertir les adjectifs de l'expression « sain et sauf » : *sauf et sain. Dans le cas des expressions verbales, certaines de ces opérations seront impossibles sans changer le sens holistique de l'expression : la passivation, la pronominalisation, le détachement, la négation, l'extraction et la relativation. Par exemple, il est impossible de mettre l'expression « ravalé sa salive » au passif sans en compromettre le sens : * la salive a été ravalée. Il n'est pas non plus possible de former une proposition relative : *la salive que Paul a ravalée, ni de faire une pronominalisation : *Paul l'a ravalée.

Cela dit, les expressions figées présentent parfois des anomalies grammaticales : on dit de ces expressions qu'elles ont une **syntaxe marquée** (Svensson 2004). Par exemple, l'expression « bel et bien » est agrammaticale : « beau » est un adjectif, tandis que « bien » est un adverbe. Pourtant, les deux mots sont liés par la conjonction « et », alors que cette conjonction relie habituellement deux mots de même nature.

2.2.4 Caractéristiques lexicales

Un autre indice permettant de vérifier si l'on est en présence d'une expression figée est le **blocage lexical** (Svensson 2004), aussi appelé « blocage des

paradigmes synonymiques » (Gross 1996), « commutation restreinte » (Forsberg 2006) et « *restricted exchangeability* » (Erman et Warren 2000). Les expressions figées contiennent au moins un mot qui ne permet aucune variation paradigmatique : la substitution de ce mot par un synonyme aurait pour conséquence de dénaturer l'expression figée. Prenons l'exemple de l'expression « c'est monnaie courante ». Si on remplace le mot « monnaie » par « devise », le caractère idiomatique se perd : **c'est devise courante*. On obtient un résultat similaire si l'on substitue l'adjectif « courante » par « fréquente » : **c'est monnaie fréquente*. Dans l'expression « en perte de vitesse », employée pour exprimer un état de dégradation, on ne pourrait pas remplacer « vitesse » par « rapidité » ou « vélocité » sans altérer l'idiomaticité de l'expression : il s'agit donc effectivement d'une EF. Les expressions figées ont un sens holistique ; si l'on change une seule de leurs composantes, elles perdent leur statut d'expressions figées et ne peuvent alors être interprétées que littéralement.

On peut également détecter un figement grâce aux mots à **contexte unique** (Svensson 2004). En effet, certains mots (archaïques, pour la plupart) sont employés exclusivement dans les expressions figées. C'est notamment le cas du mot *prou*, qui n'existe que dans la locution adverbiale *peu ou prou* (Le Petit Robert 2007), ou du mot *lurette*, que l'on ne retrouve que dans l'expression *depuis belle lurette*. Donc, lorsqu'on remarque la présence d'un mot très rare, il s'agit de vérifier si ce mot existe dans d'autres contextes, ou s'il apparaît toujours dans la même expression.

2.2.5 Caractéristiques sémantiques

Beaucoup d'expressions présentent un figement au plan sémantique : ces expressions sont dites **non compositionnelles** (Wray 2002; Svensson 2004). Voici la définition de Gross (1996) du concept de compositionnalité : « une construction donnée est dite compositionnelle quand on peut déduire son sens de celui de ses éléments composants reliés par une relation syntaxique spécifique. La non compositionnalité du sens va de pair avec des restrictions syntaxiques » (Gross 1996, p.54).

Les collocations ont généralement un sens compositionnel. Si nous reprenons l'exemple des collocations formées à partir du mot « eau », nous pouvons constater que le sens peut être déduit à partir du sens de chacun des mots :

Eau potable = [eau : « liquide transparent »] + [potable : « propre à la consommation »]

Les expressions idiomatiques, au contraire, ont toujours un sens non compositionnel, étant donné leur aspect métaphorique. Par exemple, « jeter (mettre, verser) de l'huile sur le feu », employée au sens de « envenimer une querelle » (Rey et Chantreau 2003), a un sens holistique et figuré : elle ne peut pas être interprétée par l'addition du sens de chacun des mots.

Les expressions figées, ainsi que certaines locutions et séquences préfabriquées, ont un sens non compositionnel. Par exemple, l'expression figée *entre chien et loup* est non compositionnelle, parce qu'on ne peut en déduire le sens global à partir du sens de chacun des mots qui la composent : [entre] + [chien] + [loup] ne signifie pas « au crépuscule, quand la nuit commence à tomber » (Petit Robert 2007). La non compositionnalité est étroitement liée à l'**opacité sémantique** : lorsque le sens de l'un ou de plusieurs éléments lexicaux d'une expression ne contribue pas au sens global de l'expression, cette dernière est opaque. La non compositionnalité et l'opacité sémantique sont généralement employées comme synonymes.

Il est également à noter que certaines expressions présentent un illogisme sémantique. Wray (2002) donne l'exemple des expressions anglaises « *I could care less* » et « *I couldn't care less* », qui semblent signifier des idées opposées, mais qui sont pourtant synonymes. En français, on observe le même phénomène entre les expressions « c'est le moindre de mes soucis » et « ce n'est pas le moindre de mes soucis ».

Le degré d'**opacité sémantique** aurait un effet sur la compréhension des expressions idiomatiques chez les enfants. Plusieurs études ont révélé que la compréhension des expressions transparentes était plus facile que la compréhension des expressions opaques (Nippold et Rudzinski 1993; Nippold et

Taylor 1995; Iralde et Lainé 2004; Gibbs 1987). Iralde et Lainé (2004) ont mené une recherche auprès d'enfants et d'adolescents locuteurs natifs du français, âgés de 9 à 14 ans. Les participants ont été soumis à deux tâches : la première consistait à paraphraser oralement des expressions idiomatiques ; la deuxième consistait à faire un choix forcé parmi plusieurs significations possibles. Il y avait des expressions opaques et des expressions transparentes (le classement a été fait selon le jugement préalable d'étudiants des 2^e et 3^e cycles universitaires, d'élèves du secondaire et de jeunes adultes). Les réponses des élèves étaient ensuite évaluées selon une échelle qualitative : « réponses idiomatiques », « réponses sens figuré approché », « réponses littérales », « autre sens » et « non réponse ». Bien que la transparence des expressions ne semble pas avoir eu d'effet considérable sur la capacité des élèves d'expliquer le sens idiomatique des expressions, il y a eu une distribution différentielle entre les réponses à sens figuré rapproché et les réponses littérales. Les chercheurs soutiennent donc l'hypothèse que le degré de transparence aurait une influence sur la compréhension.

L'opacité sémantique d'une expression non familière peut poser des difficultés de compréhension à un locuteur, mais ce dernier peut arriver à l'interpréter correctement en se fiant au **contexte**. L'étude de Laval (2003) menée auprès de LN du français a montré que le contexte jouait un rôle important dans la compréhension des expressions idiomatiques non familières chez les enfants 6 et 9 ans et même chez les adultes. Le contexte a également eu un effet significatif sur la compréhension des expressions idiomatiques chez les enfants dans la recherche de Levorato et Cacciari (1992), dont nous avons parlé précédemment. Cependant, d'autres études suggèrent que le contexte aide à la compréhension des expressions idiomatiques seulement jusqu'à l'âge 9 ans, mais qu'il ne serait plus nécessaire à partir de l'âge de 10 ans (Ackerman 1982). Dans notre recherche, les participants sont appelés à compléter des expressions : le contexte est essentiel, car il doit fournir suffisamment d'indices pour amener l'élève à trouver le ou les mots manquants. Par conséquent, dans notre instrument de mesure, les expressions à compléter ont été insérées dans un contexte approprié (chaque item était constitué d'une phrase de mise en contexte suivi d'une phrase lacunaire comprenant une

expression à compléter). Des explications complémentaires seront fournies dans la *Méthodologie*.

2.2.6 Caractéristiques phonologiques

Comme nous l'avons expliqué précédemment, de nombreux chercheurs (dont Wray 2002; 2007) postulent que les expressions figées sont mémorisées et emmagasinées dans le lexique mental des locuteurs comme des unités insécables. Les EF demeurent à la disposition du locuteur, qui les emploie pour communiquer (c.-à-d. pour comprendre et aussi pour s'exprimer). Donc, au lieu de créer des énoncés à partir de combinaisons libres, le locuteur a recours aux EF qui, étant stockées telles quelles dans la mémoire, ne sollicitent pratiquement aucun effort de formulation. Cette réduction de l'effort de production peut se manifester dans divers aspects de la forme phonologique d'une EF, notamment, la fluidité verbale (*fluency*), l'intonation, l'articulation, ainsi que les liaisons en français. En ce qui concerne la **fluidité**, il semble que plus une EF est répétée, plus le temps d'articulation sera écourté (Bybee 2002). En effet, la production de mots et d'expressions repose sur des comportements neuromoteurs. Plus ces comportements (c.-à-d. des séquences de mots) sont répétés, plus l'articulation devient automatique et gagne en fluidité (Bybee 2002).

L'**intonation** et l'**articulation** des expressions figées peuvent également permettre d'identifier les expressions figées dans un corpus oral. Notamment, une articulation relâchée est l'une des caractéristiques susceptibles d'indiquer la présence d'une EF. Par exemple, l'un des items du test utilisé lors de la collecte de données, « flambant neuf », a généré quelques erreurs de transcription chez les élèves. Parmi les réponses obtenues, on a trouvé « *flamant neuf » et « *flambà neuf ». Ces erreurs pourraient s'expliquer par la prononciation souvent relâchée de cette expression (quoique des troubles de perception phonologique puissent également entrer en jeu).

Parmi les autres caractéristiques permettant de retracer les EF dans la langue parlée, Wray (2002) mentionne également l'**impossibilité de marquer une pause** ou d'adopter une **intonation inhabituelle** lors de l'énonciation d'une EF. Quant aux

marques de **liaison** en français parlé (c.-à-d. la prononciation de la consonne finale d'un mot précédant un autre mot commençant par une voyelle), elles peuvent aussi se révéler un indicateur de la présence d'une EF. En effet, plus une combinaison de mots est fréquente, plus les locuteurs ont tendance à marquer la liaison. Par exemple, dans l'expression « faits et gestes », le « s » de « faits » sera audible à l'oral (du moins, lorsqu'elle est produite par des locuteurs natifs, ces derniers ayant l'habitude d'entendre cette expression prononcée ainsi). Dans le cas d'une combinaison libre (Ex. : les tulipes et les jonquilles), la consonne finale demeure habituellement muette.

Enfin, Gross (1996) fait remarquer que la répétition de sons à l'intérieur des expressions figées est un phénomène fréquent. Notez, par exemple, l'allitération dans les expressions « bel et bien » et « peu ou prou », et la rime dans l'expression « au fur et à mesure ».

2.2.7 Le figement : un phénomène relatif

Il est important de noter que la notion de figement n'est pas absolue : certaines expressions sont très fixes tandis que d'autres sont relativement flexibles. Gross (1996) parle d'ailleurs de « degré de figement » (p.6), et signale que « les suites totalement figées sont très minoritaires par rapport à celles qui ont des restrictions partielles » (Gross 1996, p. 22). Wray (2002) reconnaît elle aussi qu'il existe non seulement des expressions fixes, mais aussi des expressions semi-fixes, qui permettent une flexibilité soit au plan lexical, soit au plan grammatical.

Examinons d'abord les expressions qui sont flexibles au niveau lexical (ou paradigmatisque), c'est-à-dire que l'un des mots de l'expression peut être remplacé par un autre mot équivalent. Quand les paradigmes possibles sont soumis à des contraintes sémantiques, on parle de « variation restreinte » (*restricted variability*) ; quand il n'y a pas ou peu de restrictions, on parle de « choix ouvert » (*open slot*). Par exemple, l'expression « aux quatre coins de... » est une expression « à variation restreinte » (*restricted variability*), parce qu'on peut compléter cette expression avec différents noms (du monde/de la ville/du pays, etc). Voici d'autres exemples d'expressions à variation restreinte : à petite/grande/échelle ; à l'aube du

vingtième siècle/millénaire, il y a de l'électricité/de la fièvre/de l'amour dans l'air, à dix kilomètres/cent kilomètres/mille lieues/ à la ronde, etc. Quant aux expressions figées à choix ouvert (*open slot*), en voici quelques exemples : mettre *qqn* au défi de *qqch*, faire le point sur *qqch*, mettre *qqch* en évidence.

D'autres expressions semi-fixes varient plutôt au plan syntaxique, et admettent certaines opérations grammaticales. Par exemple, certaines EF lexicales sont « extensibles » ou « réductibles », en ce sens qu'elles comportent des composantes facultatives : *d'une (seule) traite*, *seul maître à bord (après Dieu)*, *(au fur et) à mesure que*. Un autre exemple de variation syntaxique est celui de l'expression « plus souvent qu'à son tour », où « son » peut être remplacé par un autre pronom : (mon, ton, notre, votre, leur), tandis que dans l'expression « n'avoir qu'à bien se tenir », le verbe avoir peut être conjugué à plusieurs personnes, temps et modes (il n'a qu'à bien se tenir, elles n'auront qu'à bien se tenir, etc.). En somme, il faut retenir que les expressions figées se situent généralement sur un continuum entre un figement partiel et un figement total.

2.3 La place du langage figé dans l'acquisition des langues

Nous avons jusqu'ici décrit les expressions qui font partie intégrante de l'usage courant chez les locuteurs d'une communauté linguistique. Or, nous n'avons pas parlé de la façon dont les locuteurs natifs et non natifs apprennent et utilisent ces expressions tout au long du développement de leurs compétences langagières. Effectivement, les recherches en linguistique acquisitionnelle ont démontré que les différentes formes figées occupaient une place très importante dans la communication, et ce, à toutes les étapes de l'acquisition en L1 et en L2. Puisque notre recherche porte sur des enfants locuteurs natifs et locuteurs non natifs, nous allons expliquer comment ces deux types d'apprenants perçoivent, apprennent et produisent du langage figé tout au long de leur apprentissage de la langue cible.

2.3.1 Le langage figé et l'acquisition de la langue première

Dès leur plus âge, les enfants sont exposés quotidiennement à leur langue maternelle et comprennent très tôt l'importance d'apprendre à parler pour communiquer. Comme le fait remarquer Wray (2002), les très jeunes enfants découvrent rapidement que le langage constitue un moyen d'attirer l'attention et d'obtenir ce qu'ils veulent, qu'il s'agisse de combler leurs besoins physiques ou affectifs. Les enfants sont donc hautement motivés à apprendre leur langue et à l'utiliser. Cet apprentissage commence par l'observation des sons de la langue pour en détecter les régularités phonologiques : en effet, la langue est perçue par l'enfant comme une suite de sons qu'il doit analyser afin d'en extraire les unités linguistiques significatives, c'est-à-dire les mots et les expressions, que Peters (1983) appelle des unités d'acquisition (*units of acquisition*). Les premières unités linguistiques produites par un enfant peuvent comprendre des mots isolés (ex. : papa, maman) et des formules simples (ex. : fais dodo, donne lait). Cela dit, il est important de souligner que ces unités langagières sont simplement des mots, des phrases ou des expressions apprises par cœur, qui permettent à l'enfant de s'exprimer. Certaines correspondent aux expressions conventionnelles que partagent les locuteurs adultes (*good morning, how are you, how do you do, thank you*, exemples empruntés à Peters 1983, p.11), mais pas toutes (ex. *I see you, open the door* : Peters 1983, p. 11-12.)

Signalons qu'à cette étape de l'acquisition, ces expressions peuvent être employées de façon idiosyncrasique : elles sont typiquement employées toujours sous la même forme, et sont parfois produites dans un contexte inapproprié (Peters 1983; Clark 1974; 1977). Elles peuvent également présenter une structure grammaticale plus complexe que les autres énoncés produits par l'enfant (Peters 1983; Clark 1974; 1977). Ce phénomène peut être attribuable au fait que l'enfant n'a pas encore les habiletés linguistiques et cognitives nécessaires pour analyser la structure des énoncés. Donc, avant l'âge de 20-24 mois environ, l'enfant communiquera surtout à l'aide d'expressions non analysées (Wray 2002). Ce n'est que vers l'âge de 20 à 30 mois que la grammaire commence à se former et que l'enfant commence à prendre conscience du potentiel combinatoire des unités

langagières (Wray 2002). À cet âge, l'enfant est à un stade de développement cognitif qui lui permet de « décomposer » les unités langagières jusqu'alors non-analysées, afin d'en extraire le lexique et les règles de grammaire, et de créer de nouveaux énoncés.

Or, les unités linguistiques produites par le jeune enfant, qui n'a aucune connaissance préalable du langage, ne correspondent pas nécessairement à la segmentation réelle de la langue : « since the child does not know the language, it is unreasonable to assume that the first units she or he extracts will coincide exactly with the words and morphemes of the system » (Peters 1983, p. 5). Donc, l'enfant ne réussit rarement du premier coup à segmenter correctement les unités langagières : les erreurs de segmentation sont fréquentes. Peters (1983) donne l'exemple d'un enfant ayant fait une erreur de segmentation du verbe « *behave* », qu'il a perçu comme une expression formée du verbe « *be* » et d'un attribut (*be/heyv*). Donc, lorsqu'il a tenté de conjuguer le verbe « *behave* », il a produit l'énoncé « *I am /heyv* »/ (au lieu de la forme correcte « *I am behaving* ») (p. 43). Selon Peters (1983), la segmentation sera mieux réussie si certaines conditions sont remplies. D'abord, il doit y avoir une saillance phonologique, soit à la première ou à la dernière syllabe, soit sur l'accent tonique, soit dans le rythme ou dans l'intonation. Cette saillance phonologique aide l'enfant à repérer les limites entre les unités et lui permet de comparer les nouvelles unités avec d'autres unités déjà mémorisées. Également, la segmentation sera mieux réussie si l'enfant a été exposé à plusieurs reprises à l'une des sous-unités (mots) de la séquence, et surtout si cette exposition a eu lieu sur une courte période. Ensuite, l'enfant doit être capable d'associer un sens clair à l'une des sous-unités (mots) résultant de la segmentation. Finalement, la production d'un énoncé créé à partir d'une segmentation ne doit pas faire l'objet d'une incompréhension, d'une indifférence ou d'une moquerie (Peters 1983 p.42). Donc, pendant cette phase de développement, l'enfant adopte une approche plutôt analytique qui l'amène à décomposer les unités langagières pour en extraire des mots et en déduire des règles grammaticales. Cela dit, pendant cette phase, l'enfant aurait tendance à interpréter l'input de façon littérale, selon Wray (2002). En effet, la chercheuse fait remarquer que pendant ce stade de

développement, qui se poursuit jusqu'à l'âge de 8 ans environ, l'enfant ne maîtrise pas encore certains aspects pragmatiques du langage, comme les métaphores, l'ironie, le sarcasme et les expressions non transparentes. Comme nous le verrons plus loin, le développement cognitif joue un rôle dans la compréhension des expressions non littérales.

Cela dit, même lorsque les enfants ont atteint un stade de développement qui leur permet de formuler leurs propres phrases, une proportion importante de leurs productions continue de reposer sur l'utilisation de phrases préfabriquées. L'une des raisons souvent évoquées est que les enfants ont des capacités cognitives limitées et que l'emploi d'énoncés préfabriqués diminue considérablement l'effort de traitement : « small children are, more than anyone, handicapped by short-term processing limitations » (Peters 1983, p. 82).

Chez les enfants d'âge préscolaire, les connaissances linguistiques se limitent à la langue orale, puisqu'ils ne savent ni lire ni écrire. Or, dès la première année du primaire, les enfants apprennent à lire et à écrire. L'initiation à la littérature jouera un rôle crucial dans la perception du langage, de même que sur l'acquisition du lexique et des expressions figées. Non seulement l'enfant apprend à reconnaître à l'écrit les mots et les expressions qu'il connaissait déjà à l'oral, mais aussi prend-il connaissance des frontières graphiques entre les mots et les expressions, en plus de reconnaître l'existence des mots en tant qu'unités linguistiques indépendantes de leur référent sémantique (Wray 2002). Le contact avec l'écrit a aussi des conséquences sur l'acquisition lexicale : entre 8 et 11 ans, le répertoire lexical des enfants « se consolide et s'enrichit » grâce à une exposition répétée à la langue écrite (Lété 2004). La récurrence des mots joue un rôle important dans l'acquisition du vocabulaire : selon Lété (2004), chaque occurrence doit être rencontrée plusieurs fois à l'écrit (soit au moins 2 ou 3 fois) pour qu'il y ait une trace en mémoire. Si l'exposition à l'écrit consolide et enrichit le vocabulaire des enfants, il y a lieu de croire qu'elle a également un effet positif sur l'acquisition des expressions. Cela dit, pendant la période correspondant à l'apprentissage de la lecture et de l'écriture, l'enfant a une approche très analytique (axée sur les mots et non sur les unités linguistiques plus grandes, comme les expressions). Selon Wray (2002), ce

phénomène pourrait être attribuable au caractère « non naturel » des processus d'apprentissage de la lecture et de l'écriture, qui favoriserait une approche axée sur les mots.

Tout au long de l'enfance et de l'adolescence, le locuteur développe des **connaissances linguistiques** (lexicales et syntaxiques), **extralinguistiques** et **pragmatiques** qui lui permettent de comprendre de mieux en mieux certains aspects pragmatiques et sémantiques du langage, dont notamment les expressions idiomatiques (Iralde 2004). En effet, plus un enfant vieillit, plus il connaît de mots : son vocabulaire en expansion lui permet d'apprendre encore plus de mots et d'expressions (*Matthew Principle*). Vers l'âge de 11 ans, le vocabulaire actif d'un enfant LN du français est de l'ordre de 5000 mots, dont la plupart seraient « suffisamment consolidés en mémoire pour être bien compris et, pour la plus grande partie d'entre eux, produits à l'oral et à l'écrit » (Lété 2004)⁶. À notre connaissance, il n'existe pas d'estimation du nombre d'expressions connues à cet âge ; or nous pouvons raisonnablement penser que le répertoire lexical d'expressions augmente sensiblement d'une année scolaire à l'autre. Quant aux connaissances extralinguistiques et pragmatiques, elles permettent à l'enfant d'analyser la façon dont le langage est employé dans diverses situations, d'interpréter le sens figuré des expressions et de saisir le sens figuré des expressions, de même que les métaphores, l'ironie, le sarcasme, et ainsi de suite. Signalons également que les enfants d'âge scolaire voient leurs **aptitudes cognitives** (mise en œuvre de la mémoire de travail, utilisation de stratégies cognitives et métacognitives ; capacité d'attention et de raisonnement) progresser rapidement (Blaye et Lemaire 2007), et que ce développement favoriserait la compréhension des expressions idiomatiques (Nippold 1988). L'une des stratégies cognitives employées par le locuteur consiste à employer davantage d'expressions figées : en effet, pendant la période s'échelonnant de 8 à 17 ans environ, les jeunes

⁶ Lété (2004) fait une distinction entre le lexique, qu'il définit comme « les mots qu'un enfant « risque » de rencontrer dans son parcours scolaire », et le vocabulaire de l'enfant, c'est-à-dire « son stock lexical en réception ou en production de l'écrit » (p. 248). Selon l'auteure, cela peut expliquer pourquoi d'autres chercheurs comme Anglin (1993) estiment le vocabulaire réceptif des enfants à quelques dizaines de milliers.

locuteurs produiraient de plus en plus de langage figé, découvrant qu'il est plus facile d'employer des formes préconstruites que de créer des énoncés à neuf (Wray 2002). Ces constatations ont une implication importante dans notre recherche. En effet, les participants à notre étude sont des enfants de 3^e à 6^e année. Les recherches ont démontré que pendant cette période, les enfants développent des habiletés linguistiques, pragmatiques et cognitives, et que cette évolution leur permettait de comprendre de mieux en mieux le sens et les fonctions des expressions figées. Donc, nous nous attendons à ce que cette évolution se reflète dans notre population par une meilleure performance chez les élèves plus âgés que chez les élèves plus jeunes.

Cela dit, ce n'est que vers la fin de l'adolescence, soit vers 18 ans, que la production langagière des locuteurs atteint un juste équilibre entre le langage créatif et le langage figé (Wray 2002). Les adultes locuteurs natifs connaissent et maîtrisent alors un grand nombre d'expressions de toutes sortes (collocations, expressions idiomatiques, expressions figées, etc.), et les utilisent en grande proportion dans leur discours, la principale raison étant que l'emploi de formules préfabriquées réduit considérablement l'effort de traitement (Perkins 1999). Selon Melcuk (1993), le répertoire d'expressions d'un locuteur natif du français serait de l'ordre de plusieurs dizaines de milliers. Même si l'adulte continue normalement d'être exposé à la langue écrite tout au long de sa vie, il est probable que son répertoire d'expressions demeure stagnant au fil des années : selon Lété (2004), les locuteurs adultes atteindraient un « plafonnement lexical » vers l'âge de 20 à 25 ans.

2.3.2 Le langage figé et l'acquisition d'une langue seconde.

Nous avons vu que dans l'apprentissage de la langue première, l'âge influence beaucoup la façon dont le langage est perçu et utilisé. Il s'avère que l'âge des locuteurs a également un effet sur la perception de la langue cible et l'approche utilisée (holistique ou analytique) dans le processus d'acquisition d'une langue seconde. Or, dans le cas de l'apprentissage L2, un autre facteur important entre en jeu : le contexte d'apprentissage. Les apprenants peuvent étudier la langue cible dans différents milieux : dans le cadre d'un cours de langue seconde donné dans

leur établissement scolaire (école primaire ou secondaire, université), dans un programme d'immersion, ou dans un contexte de submersion (où l'apprenant apprend la langue cible parmi des locuteurs natifs). Le contexte d'apprentissage est crucial, car il détermine en grande partie la quantité et le type d'input auquel les apprenants sont exposés et le type d'interactions auxquelles ils sont appelés à participer.

Chez les apprenants d'âge préscolaire qui apprennent la L2 dans un contexte de submersion, (par exemple, dans une garderie où la langue d'usage est la L2), l'acquisition de la L2 se fait de façon presque identique à l'acquisition de la L1, c'est-à-dire que les apprenants emploient la langue cible dans le but premier de communiquer avec les adultes afin de subvenir à leurs besoins physiques et affectifs (par ex. : *more juice, I want to get down please* (Wray 2002 p. 153). À ce stade, les séquences préfabriquées sont donc employées à titre de stratégie de communication. Les enfants LNN apprennent des expressions pour communiquer non seulement avec les adultes, mais aussi avec les autres enfants, et ce, à des fins d'intégration sociale et culturelle. Les enfants soucieux de s'intégrer à un groupe seront portés à employer des formules propres à la langue cible et à se mêler aux enfants parlant cette langue, par exemple des chansons, des comptines, des formules de ralliement : « *lunch time !* » « *happy birthday to you!* » (Wray 2002, p. 154). Les formules toutes faites seraient également employées par les enfants comme stratégie d'apprentissage de la grammaire. En effet, plusieurs chercheurs ayant étudié la production orale de très jeunes enfants LNN ont observé que ces derniers « segmentaient » certaines expressions, c'est-à-dire qu'ils s'en servaient comme outil d'analyse de la grammaire et du lexique (comme c'est le cas dans l'acquisition L1). L'étude de Wong-Fillmore (1976), par exemple, est l'une des premières recherches consacrées à l'exploration des fonctions du langage figé dans l'acquisition d'une langue seconde chez des enfants d'âge préscolaire et ayant observé le phénomène de la segmentation. Le phénomène de la « décomposition » ou « segmentation » n'est toutefois pas reconnu par tous les chercheurs comme étant une véritable stratégie d'apprentissage. Notamment, Krashen et Scarcella (1978) reconnaissent que les expressions figées constituent une stratégie de

communication efficace en début d'apprentissage, mais croient qu'elles n'ont pas de rôle significatif dans l'appropriation des règles grammaticales.

Chez les apprenants adolescents et adultes, l'apprentissage et l'utilisation de langage figé varient beaucoup selon le contexte d'apprentissage. Lorsque l'apprentissage a lieu en salle de classe, ils ont tendance à adopter une démarche plutôt analytique, ce qui les pousse à être un peu trop « créatifs », c'est-à-dire qu'ils essaient de combiner les mots à partir de leurs connaissances grammaticales, ce qui mène parfois à des formulations non idiomatiques. Mentionnons par ailleurs que dans un cours de langue, les apprenants ne sont pas placés dans une situation authentique de communication. Ils sont donc rarement appelés à communiquer un message réel, et donc moins portés à employer des expressions figées pour s'exprimer efficacement (Wray 2002). Par ailleurs, le besoin d'exprimer leur sentiment d'appartenance à leur groupe linguistique, ainsi que la pression des pairs, peut même pousser les apprenants à éviter d'employer les EF propres à la langue cible, ou d'employer des EF non natives. Même si les apprenants auront parfois recours au langage figé pour éviter la surcharge cognitive, c'est l'approche analytique qui prévaut (Wray 2002).

Nous avons vu que les apprenants débutants d'âge préscolaire (L1 et L2) segmentaient des expressions pour créer de nouveaux énoncés. Il semble que la segmentation puisse également faire partie du processus d'apprentissage des adolescents apprenant une L2 dans un milieu scolaire. Dans leurs deux études longitudinales, Myles, Hooper et Mitchell (1998; 1999) ont suivi le développement de la production langagière de 16 élèves âgés de 11 à 14 ans apprenant le français langue étrangère au Royaume-Uni. Leurs observations étaient centrées sur les trois formules suivantes : j'aime, j'adore et j'habite. Au début de leur acquisition, les élèves employaient ces séquences telles quelles, ce qui donnait parfois lieu à des emplois idiosyncrasiques erronés : « *la Monique j'adore le tennis ». Les résultats ont démontré que les séquences apprises en début d'acquisition étaient par la suite décomposées pour créer de nouveaux énoncés. En effet, les élèves se sont mis à employer le pronom « j' » « je » et les verbes « habiter », « aimer » et « adorer » de façon productive dans d'autres constructions.

Quant aux apprenants adolescents et adultes évoluant dans un contexte « naturel », la pression sociale et leurs besoins socio-interactionnels peuvent les pousser à employer davantage d'expressions figées pour communiquer. Cependant, certaines attitudes peuvent faire obstacle à l'apprentissage, dont les stratégies d'évitement, par exemple le fait d'utiliser la L1 lorsque cela est possible, ou le fait d'éviter les situations susceptibles de poser des difficultés linguistiques (Wray 2002). Parmi les recherches menées auprès de locuteurs non natifs adultes, citons notamment de Schmitt et al. (2004), qui visait à mesurer la progression des connaissances en matière de séquences préfabriquées chez des étudiants universitaires sur une période équivalant à une session universitaire. Nous présenterons ici un compte rendu de leur recherche, et nous accorderons une importance particulière à leur méthodologie, puisqu'elle a servi de fondement pour l'élaboration de notre démarche. Les participants (94 au total) à l'étude étaient des étudiants internationaux suivant des cours préparatoires au Centre for English Language Education (CELE) à l'Université de Nottingham. Pour ces étudiants, la langue d'enseignement, c'est-à-dire l'anglais, était une langue seconde. Pour 63 d'entre eux, le chinois était la langue première ; dix autres avaient le japonais comme L1 ; les 21 autres parlaient diverses langues (non précisées). Les séquences préfabriquées sur lesquelles Schmitt et ses collaborateurs se sont penchés dans leur étude étaient des expressions faisant partie du vocabulaire employé dans le contexte universitaire, tant à l'oral qu'à l'écrit. Les séquences préfabriquées ont été extraites de deux recherches portant sur les SP : celle de Biber et al. (1999), et celle de Nattinger et DeCarrico (1992), ainsi que des manuels scolaires utilisés au CELE. Par la suite, elles ont été sélectionnées en fonction de leur fréquence et de leur utilité en contexte universitaire. En effet, elles ne devaient être ni très fréquentes, ni très rares (selon les corpus suivants : British National Corpus (BNC), CANCODE, MICASE), et devaient être considérées comme utiles à apprendre dans le cadre d'une formation universitaire (selon le jugement des professeurs consultés). Beaucoup d'entre elles étaient des structateurs de discours, par exemple : *in terms of*, *on the contrary*, *on the other hand*, etc.

À partir des séquences choisies (20 au total), les chercheurs ont mis au point une batterie de tests visant à mesurer les compétences réceptives et productives des séquences préfabriquées, de même que l'aptitude et la motivation en matière d'apprentissage des langues. Le premier test administré aux étudiants était un test de production semi-guidée. La tâche consistait à compléter des séquences préfabriquées dans deux textes à trous : un texte suivi et un dialogue. Dans chaque SP, un ou deux mots étaient tronqués. Voici un exemple de phrase à compléter tirée du test :

If we look at language learning research, there i____ no evid____ that just learning from textbooks can make you a good speaker (nothing to show that this is true).

Pour compléter les expressions correctement, les étudiants devaient se fier aux indices liés au contexte, aux lettres fournies et à la paraphrase.

À partir des mêmes textes et des mêmes séquences préfabriquées, Schmitt *et al.* ont créé un autre test pour évaluer les connaissances réceptives. Il s'agissait d'un texte à trous avec choix de réponses. Contrairement au test de compétences productives, les séquences préfabriquées du test de compétences réceptives n'étaient pas incomplètes, mais manquantes. Pour chaque espace vide, les participants devaient choisir la bonne réponse parmi cinq options. Voici un exemple de phrase à trous tirée du test :

I've been watching the news report and they say that (11)_____ the international debts of poorer countries might be cancelled.

- a. there's a good chance that
- b. it seems to be happening that
- c. the evidence is increasing that
- d. people are thinking that
- e. I DON'T KNOW

L'option « I don't know » permettait aux participants qui ne connaissaient pas la réponse, de ne pas répondre quelque chose au hasard.

En tout, l'équipe de Schmitt *et al.* a administré cinq tests aux participants (dans l'ordre de passation) :

1. Production semi-guidée des SP
2. Aptitudes en matière d'apprentissage d'une L2
3. Attitude et motivation en matière d'apprentissage d'une L2
4. Niveau de vocabulaire 3000
5. Niveau de vocabulaire 5000
6. Connaissances réceptives des SP

Les étudiants ont été soumis à la batterie de tests à deux reprises, soit une fois au début de leur session préparatoire (« presessional course »), et une seconde fois à la fin de la session (sauf le questionnaire sur les aptitudes d'apprentissage d'une L2). La performance des étudiants lors de leur première passation du test a démontré que ces derniers avaient, au préalable, une bonne connaissance des séquences préfabriquées de nature académique. En moyenne, les scores obtenus pour le test de connaissances réceptives avoisinaient les 17/20, alors que pour le test de connaissances productives, la performance moyenne s'élevait à près de 13/20.

Au cours de leur session préparatoire, les étudiants ont été exposés à chacune des vingt (20) séquences préfabriquées de la batterie de tests. En effet, chacune des SP se trouvait au moins une fois dans les manuels scolaires utilisés dans le cadre de leurs cours ; chacune a fait l'objet d'un enseignement explicite au moins une fois pendant la session. À la fin de la session, les chercheurs ont pu comparer les résultats obtenus entre le prétest et le posttest. Il s'est avéré que les étudiants ont vu leur score s'améliorer de façon significative dans le posttest. Sur le plan des connaissances réceptives, la moyenne du groupe est passée de 17/20 à 19/20, alors que sur le plan des connaissances productives, elle est passée de 13/20 à 16/20.

Dans leur analyse, Schmitt *et al.* ont également mis en corrélation les gains de performance avec les autres variables, c'est-à-dire les aptitudes, la motivation et l'attitude face à l'apprentissage de la L2. Contre toute attente, aucun de ces facteurs ne semble avoir exercé une influence significative sur les hausses de performance. Pour tenter d'expliquer ce phénomène, les chercheurs posent l'hypothèse que les

dispositions psychologiques et intellectuelles des apprenants n'aient pas un impact direct sur l'acquisition des séquences préfabriquées ; autrement dit, que leur influence serait modifiée par d'autres facteurs liés au contexte d'apprentissage.

Dans la méthodologie, nous mettrons en évidence les différences et les similarités entre la recherche de Schmitt *et al.* et la nôtre et nous expliquerons quelles ont été les adaptations nécessaires compte tenu de la population étudiée.

En somme, que ce soit en langue première ou en langue seconde, l'âge joue un rôle dans l'approche d'apprentissage adoptée et dans l'utilisation de langage figé. En langue première, la démarche d'apprentissage part généralement d'une approche holistique et va vers une approche analytique, pour finalement se stabiliser à l'âge adulte. En langue seconde, c'est plutôt l'inverse qui se produit : l'apprenant adulte part généralement d'une approche analytique, puis ira peu à peu vers une approche holistique : « where the first language learner starts with large and complex strings, and never breaks them down more than necessary, the post-childhood second language learner is starting with small units and trying to build them up » (Wray p.206).

2.3.3 L'emploi d'expressions figées pour atteindre les compétences d'un locuteur natif

Tel que mentionné précédemment, l'emploi de langage figé favoriserait la fluidité verbale (« fluency ») chez les locuteurs natifs et non natifs (Wood 2006; Pawley & Syder 2000; 1976, Raupach 1984 Towell et al. (1996). En effet, puisqu'elles sont stockées comme une seule unité lexicale, elles sollicitent beaucoup moins d'effort de traitement au moment de la production, ce qui augmente la vitesse d'élocution, diminue la longueur et la fréquence des pauses, et permet de plus longues séquences de discours ininterrompues entre les pauses (Wood 2006). Pawley et Syder (2000) adoptent d'ailleurs une position assez claire sur le rôle des expressions figées dans la fluidité verbale : « it is the store of memorized constructions and expressions, more than anything, that is the key to nativelike fluency » (Pawley et Syder 2000 p. 195). Également, il semble que l'utilisation d'EF aiderait l'apprenant à structurer son discours : certains énoncés étant préfabriqués, l'apprenant peut

accorder plus d'attention à la planification des autres énoncés (Raupach 1984). Par ailleurs, il semble qu'une utilisation appropriée des EF permettrait aux apprenants de s'exprimer de façon plus idiomatique (plus « naturelle ») dans leur langue seconde : « the formulaic sequences used by native speakers are not easy for learners to identify and master, and (...) their absence greatly contributes to learners not sounding idiomatic » (Wray 2002, p. 176). La fluidité verbale et l'idiomaticité sont des compétences linguistiques très importantes, car ce sont des habiletés que seuls les locuteurs natifs ou quasi natifs peuvent atteindre. Dans notre recherche, nous présumons que les participants allophones s'expriment déjà de façon fluide en français parce qu'ils fréquentent une école francophone depuis l'âge de cinq ans. Or, nous croyons qu'ils connaissent moins bien certaines expressions que les francophones, surtout en ce qui a trait aux habiletés productives : nous savons que les enfants doivent très bien connaître les mots pour pouvoir les utiliser (Lété 2004), donc, il y a lieu de croire que les expressions doivent également être très bien connues pour être employées par les enfants. Si les enfants allophones connaissent un moins grand nombre d'expressions figées que les enfants francophones, ils risquent d'avoir davantage de difficultés à accomplir une tâche de production guidée telle que celle qui sera utilisée dans notre étude.

2.4 Conclusion du chapitre

Dans la première partie de ce chapitre, nous avons décrit les cinq formes de langage figé, soit les collocations, les séquences préfabriquées, les expressions idiomatiques, les locutions et les expressions figées, et nous avons relevé les similarités et les différences entre ces types d'expressions. Nous avons poursuivi avec une analyse des caractéristiques communes aux différentes formes de langage figé, qui servent à la fois à décrire et à identifier les formes figées dans la langue. Nous avons également expliqué l'impact de certaines caractéristiques (fréquence, opacité/transparence) sur l'acquisition des EF. Soulignons que les concepts abordés dans cette section contribueront à la présente recherche de deux façons. D'abord, les caractéristiques propres aux expressions figées, telles qu'évoquées dans les définitions, dans les caractéristiques et dans la typologie, guideront la sélection des

expressions figées en vue de l'élaboration de l'instrument de mesure. Ensuite, l'une d'entre elles, la fréquence, servira de paramètres d'analyse lors de l'interprétation des résultats, notamment pour expliquer les taux de réussite des items. Cela dit, la discussion a ensuite été orientée vers l'apprentissage et la production de formes figées au cours du développement des compétences linguistiques chez les locuteurs natifs et non natifs. Nous avons expliqué que des facteurs individuels, comme la maturité cognitive, les connaissances linguistiques et pragmatiques, ainsi que le milieu d'apprentissage et le degré d'exposition à la langue orale et écrite pouvaient influencer les connaissances réceptives et productives des expressions. Rappelons que dans la présente recherche, nous nous intéressons entre autres à l'influence du niveau scolaire et du groupe linguistique sur les connaissances en matière d'expressions figées. Puisque l'avancement dans la scolarisation va de pair avec une augmentation de l'âge et un développement au plan cognitif, il y a lieu de croire que des enfants de différents niveaux scolaires se distinguent dans leur capacité à produire des expressions appropriées dans un contexte donné. Quant au facteur « groupe linguistique », nous posons l'hypothèse que le groupe composé d'enfants francophones performera mieux à l'épreuve de production guidée que les allophones, car grâce à leur exposition accrue au français, ils ont davantage de chances de développer des connaissances étendues et solides en matière d'expressions.

CHAPITRE III

QUESTIONS DE RECHERCHE

Tout d'abord, rappelons au lecteur que l'objectif général de cette recherche est de mesurer les connaissances d'enfants allophones et francophones de niveau primaire en matière d'expressions figées. Il s'agit maintenant de préciser les questions de recherche auxquelles nous tenterons de répondre.

- 1) Est-ce que la performance des élèves s'améliore au fur et à mesure qu'ils avancent dans le système scolaire ?

Pour répondre à cette question, nous ferons passer un test écrit à des élèves de 3^e, 4^e, 5^e et 6^e année du primaire. Les enfants de 1^{re} et de 2^e année ont été exclus de cette étude étant donné la relative complexité de la tâche à exécuter et les compétences linguistiques qu'elle requiert (c.-à-d. la capacité de lire et d'écrire de façon autonome). Cela dit, il est à noter que les niveaux scolaires évalués correspondent aux deuxième et troisième cycles du primaire (3^e et 4^e années : 2^e cycle ; 5^e et 6^e années : 3^e cycle). À partir des résultats obtenus, nous vérifierons s'il semble y avoir une progression des connaissances d'une année scolaire à l'autre. À cette étape, il s'agira d'effectuer une analyse globale incluant les résultats des allophones et des francophones.

- 2) Y a-t-il une différence significative entre les élèves allophones et francophones au plan des connaissances en matière d'expressions figées ?

Rappelons que les jeunes allophones ne profitent pas de la même exposition au français que les jeunes francophones : hors du contexte scolaire, les élèves allophones sont généralement moins en contact avec le français que les élèves francophones, surtout si la langue parlée à la maison n'est pas le français (voir *Problématique*). Puisque leur milieu familial leur offre peu (ou pas) d'input langagier en français, les enfants allophones ont moins d'occasions de se familiariser avec les expressions figées de cette langue. On peut donc s'attendre à ce que les résultats des allophones au test soient plus faibles que ceux des francophones.

- 3) Y a-t-il une différence de performance entre les filles et les garçons ?

Un rapport publié par le MEQ (2004) révèle que de façon générale, les filles performant généralement mieux que les garçons à l'école. Cet écart est particulièrement marqué en ce qui a trait à l'écriture et à la compréhension de l'écrit, tel que reflété dans les résultats obtenus aux épreuves uniques du ministère de l'Éducation. Le même rapport explique que l'écart entre la réussite scolaire des garçons et des filles est plus marqué dans les milieux socioéconomiques défavorisés. Il est donc possible que les filles se démarquent par rapport aux garçons en ce qui a trait aux expressions figées.

- 4) Quelle est l'influence des caractéristiques linguistiques des EF ciblées sur la performance des élèves par degré scolaire, par groupe linguistique et par sexe ?

Nous avons considéré la fréquence des expressions, la fréquence des mots ciblés et le nombre de concurrents des mots ciblés contenus dans les expressions (c'est-à-dire le nombre de mots plus fréquents commençant par les mêmes lettres).

Bien entendu, nombreux sont les facteurs individuels pouvant influencer sur l'acquisition des EF : l'influence de la lecture en français effectuée en dehors du contexte scolaire et, dans le cas des allophones, l'influence de la langue première et le nombre d'années écoulées depuis le début de l'apprentissage du français. En ce qui concerne l'influence de la lecture, elle offrirait sans doute une perspective intéressante à une recherche comme celle-ci. Toutefois, ce facteur est assez complexe à mesurer, car il implique de nombreux aspects (le type de matériel lu, la fréquence des lectures, le temps consacré à la lecture, etc.). Pour mesurer l'influence de ce facteur, il aurait fallu effectuer une analyse approfondie des habitudes de lecture de chaque élève, ce qui n'était pas possible étant donné le nombre de participants et le temps alloué par l'école à notre collecte de données. Quant à l'influence de la langue première (créole, mandarin, espagnol, etc.), elle n'a pas été prise en considération vu la nature exploratoire de cette recherche et la très grande variété de langues parlées par les élèves de l'école. En effet, nous avons préféré garder une approche globale et adopter un système bipartite de classement des participants (allophones/francophones). En ce qui a trait au nombre d'années écoulées depuis le début de l'apprentissage formel du français, ce facteur n'a pas pu être étudié puisque la grande majorité des participants à cette étude fréquentaient une classe régulière francophone depuis la maternelle.

Il aurait également été intéressant de mesurer l'influence de la transparence et de l'opacité sémantique sur la réussite au test. Or, puisque ces caractéristiques ne sont pas quantifiables, il nous est impossible de mesurer leur impact à l'aide de tests statistiques. Un moyen d'estimer le degré d'opacité des expressions est de faire appel à un grand nombre d'informateurs afin de minimiser la subjectivité. Cependant, les contraintes de temps ne nous ont pas permis d'effectuer cette démarche.

La performance des élèves au test fera l'objet d'une discussion dans le chapitre *Analyse et interprétation des résultats*, où nous proposerons des réponses aux

questions de recherche. Cela dit, nous verrons, dans le prochain chapitre, la démarche méthodologique adoptée pour répondre aux questions de recherche.

CHAPITRE IV

APPROCHE MÉTHODOLOGIQUE

La méthodologie adoptée dans cette étude est fondée en grande partie sur celle de Norbert Schmitt *et al* (2004), et plus particulièrement en ce qui a trait à la conception de l'instrument de mesure (leur étude a été expliquée dans le cadre théorique.) Essentiellement, la recherche de Schmitt *et al.* et la présente étude ont le même objectif, soit celui de mesurer les connaissances d'apprenants en matière d'expressions figées dans leur langue seconde. Les deux recherches concernent la langue écrite et ont été menées dans un contexte authentique d'apprentissage, ou plus précisément, en milieu scolaire. De plus, les participants aux deux études apprennent leur langue seconde dans un contexte de submersion. Toutefois, les deux recherches ne sont pas en tous points semblables. Les différences les plus importantes entre notre recherche et celle de Schmitt *et al.* se résument en deux points : la langue cible (l'anglais dans le cas de Schmitt *et al.*, le français dans la présente recherche), ainsi que l'âge (et le niveau scolaire) des participants (étudiants universitaires dans l'étude de Schmitt *et al.*, élèves du primaire dans cette étude). En effet, l'instrument de mesure de Schmitt *et al.*, conçu pour des étudiants universitaires apprenant l'anglais langue seconde, a dû être adapté aux fins de la présente recherche, à la fois en fonction des enfants allophones et des spécificités du français.

Lorsque l'on doit créer ou adapter un outil d'évaluation pour des enfants, on doit inévitablement faire face à des obstacles et à des contraintes. D'abord, le niveau de

difficulté doit être adapté relativement à la nature de la tâche, à la consigne, au vocabulaire employé et au contenu du test. Ensuite, le contenu du test doit correspondre au vécu des enfants et, autant que possible, stimuler leur intérêt. De plus, la durée du test doit être limitée, car les enfants doivent demeurer concentrés du début à la fin de l'épreuve. Enfin, dans le cas d'une collecte de données réalisée en milieu scolaire, les contraintes de temps ne sont pas non plus à négliger. Les adaptations effectuées dans notre instrument de mesure seront expliquées en détail dans la sous-section « conception de l'instrument de mesure ».

4.1 Participants

La collecte de données a eu lieu dans une école primaire francophone publique située dans un quartier multiethnique à Montréal. Il s'agit d'une école très défavorisée (classée au 10^e rang décile de l'indice de milieu socioéconomique (IMSE) et également au 10^e rang décile de l'indice du seuil de faible revenu (SFR), selon les statistiques du MELS)¹. Donc, cette école compte parmi les plus pauvres du Québec. Le test a été administré à 165 élèves (92 filles et 73 garçons) de la 3^e à la 6^e année du primaire (2^e et 3^e cycles). Il y avait au total 102 francophones et 63 allophones. Parmi les langues premières des élèves allophones, le créole et le mandarin étaient les plus fréquentes.

La répartition des participants par groupe linguistique, année scolaire et sexe est présentée dans le tableau 4.1 et l'âge des participants par groupe linguistique et par année scolaire est présenté dans le tableau 4.2.

¹ Document PDF consultable à l'adresse suivante :
http://www.mels.gouv.qc.ca/stat/Indice_defav/Documents03/Ensemble_indices_0304.pdf

Tableau 4.1 Répartition des participants selon l'année, le groupe linguistique et le sexe

Année	Francophones		Total fr.	Allophones		Total all.
	Filles	Garçons		Filles	Garçons	
3e	11	5	16	5	6	11
4e	11	8	19	6	8	14
5e	11	15	26	15	4	19
6e	23	10	33	5	8	13
DGA	2	6	8	3	3	6
Total	58	44	102	34	29	63

Tableau 4.2 Âge des participants selon l'année scolaire et le groupe linguistique

Année	Francophones			Allophones		
	Âge	Âge	Âge	Âge	Âge	Âge
	min.	max	moyen	min.	max	moyen
3e	8,68	10,47	9,37	8,59	10,16	9,17
4e	9,66	11,46	10,36	9,48	11,04	10,33
5e	10,65	12,53	11,19	10,62	11,90	11,26
6e	11,79	12,91	12,32	11,64	13,21	12,16
DGA	11,69	13,24	12,26	11,87	12,68	12,37

Afin d'éviter toute ambiguïté, il est nécessaire de préciser ce que l'on entend par « allophone » et « francophone ». Rappelons à cet effet la définition d'allophone présentée dans la *Problématique* : « personne dont la langue d'usage (celle qui est parlée à la maison) ou la langue maternelle est différente de la ou des langues officielles du pays où elle se trouve » (Source : GDT). Il a déjà été question, dans la problématique, du statut du français au Québec et du Canada. Il a été établi que le

français était la seule langue officielle du Québec, et que puisque cette recherche avait lieu sur le territoire québécois et dans une école francophone, il était convenable de définir les participants qui parlaient une autre langue que le français à la maison comme des allophones².

Afin de connaître la langue que parlaient les élèves à la maison, on a demandé à ces derniers de répondre à un questionnaire de renseignements concernant les langues parlées le plus souvent avec leurs parents, avec leurs frères et sœurs, ainsi qu'avec leurs amis. Toutefois, après réflexion, il a été décidé que c'est la langue parlée avec les parents qui déterminerait le statut linguistique des élèves. La raison en est que ce sont les parents qui transmettent leur langue première à leurs enfants, et il y a lieu de croire que dans bien des cas, cette langue soit maintenue dans les interactions parent-enfant au fil des années, même si les enfants sont scolarisés dans une autre langue (c'est du moins ce que laissent croire les renseignements donnés par les élèves, qui indiquent la dominance d'une langue étrangère à la maison chez 63 élèves sur 165 (38 %) et la codominance du français et d'une autre langue à la maison chez 43 élèves sur 165 (26 %)).

Par conséquent, les élèves ayant affirmé parler surtout (ou seulement) le français avec leurs parents ont été considérés comme des francophones, tandis que ceux qui ont déclaré parler surtout (ou seulement) une autre langue que le français avec leurs parents ont été considérés comme allophones. Quant aux élèves qui ont déclaré parler français aussi souvent qu'une autre langue avec leurs parents, ils ont également été considérés comme des francophones.

Il est vrai que ce seul critère d'identification comporte des limites, et que le français est parfois adopté comme langue d'usage dans des familles dont la langue

² Étant donné la nature de cette étude, il n'était pas pertinent de faire une distinction entre allophones et anglophones. Les rares élèves anglophones ont été considérés comme des allophones, puisque leur langue première n'est pas la même que la langue officielle du Québec.

première n'est pourtant pas le français³. Toutefois, il fallait trouver une façon d'identifier systématiquement les groupes linguistiques, et ce moyen semblait à la fois le plus simple et le plus significatif. En effet, la langue parlée avec les frères et les sœurs aurait, elle aussi, pu être considérée comme un indicateur de la langue première, puisqu'elle est présente dans le contexte familial. Toutefois, cette variable apportait son lot d'inconvénients, surtout dans le cas des enfants uniques. Par ailleurs, il nous semblait que la langue dominant les échanges entre frères et sœurs était peut-être davantage susceptible de varier en fonction du contexte (à la maison devant les parents, à la maison en l'absence des parents, entre amis, à l'école, dans les activités parascolaires, etc.) et ces variables auraient complexifié les données. D'autres variables, comme le nombre et l'âge des frères et des sœurs, auraient alors dû être prises en considération. Quant à la langue parlée avec les amis, elle devait servir à indiquer si l'enfant communiquait surtout en français ou en une autre langue dans son réseau social hors des contextes scolaire et familial. Encore une fois, la prise en compte de cette variable impliquerait de nombreux détails complexes à mesurer et qui par ailleurs seraient difficiles à laisser au jugement des enfants (par exemple : nombre d'amis parlant français/autre langue, nombre d'heures passées hebdomadairement avec les amis, type d'activités réalisées avec les amis, etc.).

4.2 Choix de l'instrument de mesure

Tel que mentionné précédemment, notre instrument de mesure est inspiré du test de production guidée mis au point par Norbert Schmitt *et al.* (2004) dans le cadre de leur étude longitudinale sur la connaissance et l'acquisition des expressions figées.

³ À Montréal, sur les 760 445 citoyens qui se déclarent allophones, 32 450 parlent français en plus d'une langue non officielle à la maison, contre 18 945 qui parlent anglais en plus d'une langue non officielle à la maison. On dénombre 6920 personnes qui parlent français et anglais en plus d'une langue non officielle. (Source : Statistique Canada, Recensement de la population de 2006, Produit no 97-555-XCB2006028 au catalogue de Statistique Canada).

Plusieurs raisons nous ont amenée à choisir d'utiliser un test de production guidée comme outil de mesure. D'abord, étant donné nos questions de recherche, il était préférable que les réponses attendues soient les mêmes pour tous les participants et que l'évaluation soit objective et quantifiable. Ensuite, le temps alloué à la passation du test étant assez limité, il fallait créer un test auquel les élèves pourraient répondre rapidement et spontanément. De plus, étant donné le nombre limité d'items dans le test, il était préférable d'éviter un type de test qui aurait permis aux élèves de répondre au hasard. Le test de Schmitt *et al.*, en plus de correspondre à tous ces critères, avait donné des résultats significatifs dans leur recherche. De surcroît, une tâche consistant à compléter des mots tronqués nous paraissait intéressante et novatrice. C'est pourquoi nous nous sommes servie du test de Schmitt *et al.* comme modèle pour la conception de notre instrument de mesure.

4.3 Choix des livres pour l'extraction des expressions figées

Les expressions ayant servi à la conception du test devaient correspondre à un niveau de difficulté adapté à des élèves de 2^e et 3^e cycles du primaire, autrement dit les 3^e, 4^e, 5^e et 6^e années. C'est pourquoi toutes les expressions sélectionnées proviennent de livres pour enfants, soit des manuels scolaires et des romans pour enfants et adolescents. Les manuels scolaires consultés (17 au total) étaient de matières variées, soit l'histoire, la géographie, les mathématiques, les sciences naturelles et le français, et étaient destinés à des élèves des 2^e et 3^e cycles du primaire, et du 1^{er} cycle du secondaire (secondaire 1, 2 et 3). Tous les manuels étaient relativement récents ; les éditions dataient d'au plus dix (10) ans.

D'autres expressions ont été extraites de la littérature jeunesse ; les livres choisis étaient des romans publiés aux éditions *La courte échelle*, une maison d'édition spécialisée en littérature jeunesse. Le niveau de difficulté des livres a pu être contrôlé grâce au classement proposé sur le site de l'éditeur.⁴ Les livres sélectionnés faisaient partie soit de la collection *Jeunesse*, dont les titres sont

⁴ Adresse du site de la maison d'édition: <http://www.courteechelle.com/>
Site consulté en mars 2007 et en novembre 2008.

destinés aux enfants de 9 à 11 ans, ou de la collection *Roman +*, qui s'adresse aux enfants de 12 ans et plus. Aux éditions *La courte échelle*, les livres sont classés dans l'une ou l'autre des collections selon les critères suivants : la thématique, le niveau de difficulté et le registre de langue, ainsi que la longueur du manuscrit.⁵

Il est important de noter que les expressions figées ont été extraites de romans jeunesse et de manuels scolaires dans le but de cibler des expressions que les élèves de 2^e et 3^e cycles sont susceptibles de rencontrer dans les ouvrages qui leur sont destinés. Donc, les participants à notre étude n'avaient pas nécessairement déjà lu ou utilisé les ouvrages que nous avons choisis pour notre recherche.

4.4 Description des expressions entrant dans la composition du test

Bien entendu, toutes les propriétés abordées dans le cadre théorique ne s'appliquent pas à tous les types de recherche. Dans le cas d'une étude évaluative comme celle-ci, les expressions choisies doivent être d'un niveau de difficulté approprié pour les participants (élèves de 2^e et 3^e cycles du primaire, dont une forte proportion d'allophones). C'est pourquoi, dans cette recherche, la provenance, la longueur et le niveau de difficulté des EF ont été contrôlés. Rappelons la définition opérationnelle que nous avons élaborée aux fins de notre recherche : *une expression figée est une suite d'au moins deux mots graphiques et occupe, à même titre que les mots simples, une fonction grammaticale précise au niveau de la phrase. Elle est caractérisée par un figement, partiel ou total, soit au plan syntaxique, lexical ou sémantique. Elle fait partie intégrante du vocabulaire d'usage des locuteurs natifs d'une langue.*

Conformément à cette définition, seules les expressions correspondant aux critères suivants étaient susceptibles d'entrer dans la composition du test :

⁵ La collection *Roman +*, qui était publiée dans les années 1990, a depuis été subdivisée en quatre collections : collection *Ado* (12 à 14 ans), collection *Ado +* (14 à 16 ans), collection *Jeune adulte* (16 ans et plus) et collection *Poésie* (12 ans et plus). Quant aux critères de classement des manuscrits, ils ont été obtenus par le biais d'une communication téléphonique avec l'éditeur, le 14 décembre 2007.

1) Expressions de deux mots ou plus (incluant les mots grammaticaux : prépositions, déterminants, conjonctions). En effet, selon les définitions de Wray (2002) et de Forsberg (2006), le nombre minimal de mots formant une EF est fixé à deux. Étant donné la relative difficulté de la tâche, le nombre de mots de chaque expression du test ne dépassait pas cinq (incluant les mots grammaticaux).

2) Elles exercent une fonction syntaxique au niveau de la phrase (elles fonctionnent soit comme un nom, un adverbe, un verbe, un adjectif, une préposition ou une conjonction.)

3) Elles présentent un figement partiel ou total soit au plan syntaxique, lexical, ou sémantique. La présence de ce figement est démontrable par le biais de diverses opérations transformationnelles indiquant un blocage (par exemple, l'inversion, la commutation, la pluralisation, etc.)

4) Elles font partie des conventions linguistiques des locuteurs natifs. Pour confirmer ce critère, nous avons vérifié si les expressions extraites étaient consignées comme des expressions ou des locutions soit dans *Le Robert des expressions et locutions*, soit dans *Le Petit Robert* ou encore dans *Le Multidictionnaire* de la langue française. Puisque l'extraction des expressions figées était initialement fondée notre intuition, ce critère a permis de certifier que les expressions sélectionnées étaient officiellement admises comme des expressions figées.

5) Les expressions figées propositionnelles (celles qui constituent des phrases à part entière) ont été exclues, puisque contrairement aux autres types d'expressions, elles n'exercent pas de fonction grammaticale au niveau de la phrase.

4.5 Extraction des expressions figées

Au fil de nos lectures, nous prélevions systématiquement toutes les expressions qui semblaient constituer des expressions figées. Le tri allait être effectué ultérieurement, selon les critères de sélection préétablis.

Soulignons qu'à cette étape de notre recherche, le recours à notre intuition de locutrice native s'est révélé inévitable. En pratique, il était impossible de chercher

systématiquement chaque syntagme dans le dictionnaire, au cas où celui-ci formerait une expression figée : c'est pourquoi nous notons tous les syntagmes qui nous étaient familiers, c'est-à-dire ceux que nous avons déjà lus ou entendus dans d'autres contextes. Bien que cette méthode puisse paraître discutable à première vue, elle se révèle souvent une étape incontournable dans les recherches portant sur les expressions figées : « it is probably true to say that intuition plays a role in the identification of formulaic sequences in almost all studies, including frequency-based ones » (Wray 2007, p. 151). Bien sûr, le simple fait que le recours à l'intuition soit courant dans ce type de recherche ne justifie pas nécessairement cette pratique. Pourtant, il semble que cette méthode soit généralement acceptée comme étant un bon point de départ : « for many linguists intuition is the legitimate basis of sound research » (Wray 2007, p.151). Dans tous les cas, les chercheurs doivent faire preuve de discernement : « the important distinction here is between the use of introspection as data and the essential utilisation of our intuitions to form hypotheses and analyse data » (Butler 2004, p.150). Le recours à l'intuition comme guide dans une recherche sur les expressions figées est donc souvent nécessaire, comme ce fut le cas pour la présente recherche.

4.6 Sélection des expressions figées

Au total, 418 expressions différentes ont été extraites des manuels scolaires et des romans. Certaines d'entre elles, étant donné leur caractère métaphorique saillant (par ex. ; *chercher une aiguille dans une botte de foin*, *grimper dans les rideaux*, *tirer le diable par la queue*), n'ont pas été retenues puisqu'elles ont été considérées comme des expressions idiomatiques. D'autres expressions, dont le degré de fixité était plus faible, semblaient appartenir davantage aux collocations (par ex. : *subvenir à leurs besoins* ; *sur une base volontaire*). Enfin, certaines expressions, qui étaient trop longues en termes de nombre de mots (par ex. : *penser tout haut ce que les autres pensent tout bas*) ou qui n'apparaissaient dans aucun des dictionnaires précités, ont été exclues. Parmi les expressions qui répondaient à tous les critères de sélection, dix-huit (18) ont été retenues pour la création de l'instrument de mesure. Ces expressions sélectionnées sont celles qui se prêtaient

le mieux à l'instrument de mesure choisi (expressions à compléter), et qui étaient assez facilement reconnaissables pour être complétées à partir d'un contexte limité (puisque le contexte entourant chaque expression était de deux phrases seulement). Nous avons également privilégié la variété des types d'expressions (adverbiales, adjectivales, etc.), afin que les expressions tronquées n'occupent pas toujours la même fonction dans la phrase. Voici les expressions retenues, en ordre d'apparition sur le test :

À l'occasion de
 Tout le monde
 Plus ou moins
 En ce qui concerne
 Être à la hauteur
 Bel et bien
 Tout au long
 Aux prises avec
 De manière à
 Au fur et à mesure
 À tour de rôle
 À toute vitesse
 Depuis belle lurette
 À portée de la main
 Les quatre coins du monde
 Sain et sauf
 Flambant neuf
 À perte de vue

4.7 Conception de l'instrument de mesure

Chaque item du test consistait en une phrase de mise en contexte, suivie d'une phrase lacunaire. La tâche des élèves consistait à compléter les phrases lacunaires avec l'expression recherchée en se fiant au contexte pragmatique et sémantique. Dans chaque phrase lacunaire, il y avait soit un, soit deux mots tronqués à compléter. Pour donner un indice de l'expression recherchée, les éléments suivants étaient fournis :

- 1) Les mots grammaticaux d'une ou deux lettres (articles, prépositions);
- 2) La ou les premières lettres de chaque mot contenant trois lettres ou plus.

Les indices fournis par le contexte ainsi que les premières lettres de chaque mot restreignaient les possibilités, de sorte qu'une seule réponse n'était possible pour chaque élément du test.

Voici un exemple d'item tiré du test :

Les prochains Jeux Olympiques auront lieu à Pékin en 2008. En vue de cette grande compétition, les athlètes doivent s'entraîner plusieurs heures par semaine pour être à la hau_____.

Mis à part les deux (2) exemples présentés au début du test, l'instrument de mesure comportait dix-huit (18) phrases lacunaires, chacune comportant une seule EF à compléter (avec soit un, soit deux mots tronqués à trouver).

4.7.1 Coupure des mots

Bien que Schmitt *et al.* aient bien détaillé leur démarche méthodologique, la procédure employée pour la coupure des mots n'a pas été précisée dans leur article. Néanmoins, en examinant leur test, on peut dégager certaines tendances, notamment que le nombre de lettres fournies semble généralement proportionnel à la longueur des mots. Cela dit, puisque le test de Schmitt *et al.* s'adressait à des adultes de niveau universitaire et que la langue cible était l'anglais, les contraintes phonétiques et syllabiques n'étaient pas les mêmes que pour la présente recherche.

Pour notre part, nous avons déterminé le nombre de lettres fournies en fonction du nombre de syllabes et des particularités phonétiques de chaque mot tronqué. En règle générale, la coupure des mots a été effectuée selon la procédure suivante :

- Pour les mots à une syllabe : après la première lettre

- Pour les mots à deux syllabes ou plus : après la première syllabe

Ce choix a été fait principalement pour respecter la phonétique du français. Effectivement, en français, la prononciation d'une lettre donnée dépend souvent des lettres qui la précèdent et qui la suivent. Or, dans une tâche consistant à compléter des mots tronqués, les lettres fournies donnent un indice phonétique crucial pour trouver les lettres manquantes, d'autant plus que beaucoup de participants étaient des enfants allophones. C'est pourquoi il a fallu porter une attention particulière à la façon dont se prononçaient les syllabes fournies : leur prononciation devait correspondre à leur prononciation réelle à l'intérieur du mot entier.

Cependant, pour certaines expressions figées du test, les coupures entre les lettres fournies et les lettres à ajouter ont fait exception à la règle. Nous expliquerons ici pourquoi certains items ont dérogé à la règle.

1- à l'occa_____ de : si l'on avait fourni seulement « oc » au lieu de « occa », les élèves n'auraient pas su si le mot commençait par le son [ok], [os] ou [o ouvert k], [o ouvert s].

8- aux pri_____ av_____ : dans ce cas-ci, c'est le mot « avec » qui déroge à la règle : normalement, nous n'aurions fourni que la première syllabe, c'est-à-dire [a]. Cependant, nous avons jugé préférable de donner [av], car le [a] seul aurait pu être confondu avec le verbe avoir ou avec la préposition « à ».

9- au fur et à m_____ : même si le mot « mesure » aurait normalement dû être coupé après la première syllabe, seule la première lettre a été donnée. Dans ce cas, c'est le caractère unique du mot « fur » qui nous a incitée à prendre cette décision : effectivement, en français moderne, ce mot n'existe que dans l'expression « au fur et à mesure ».

15- qua_____ coins du m_____ : seule la lettre initiale « m » a été fournie, car le mot « monde » apparaissait déjà au début du test. La syllabe « mon » aurait donc constitué un indice trop évident pour les participants qui ne connaissaient pas la réponse.

17- fl_____ neuf : la syllabe « flam », écrite seule, se prononce [flam], alors que dans le mot « flambant », elle se prononce plutôt [flā].

En somme, la coupure des mots a été faite en respectant les particularités du français.

4.7.2 Adaptations effectuées en fonction des participants

Essentiellement, les items de notre test ont été construits de la même façon que ceux de Schmitt *et al.*, à quelques différences près. En effet, notre test étant destiné à des élèves allophones de niveau primaire, quelques adaptations ont été jugées nécessaires. La première adaptation concernait le type d'expressions figées. L'équipe de Schmitt *et al.* a choisi de se concentrer sur les expressions figées de nature académique, c'est-à-dire des expressions fréquentes, et donc utiles, dans un contexte universitaire. Ainsi, la plupart des expressions figées à l'étude étaient des structateurs de discours (ex. : *as far as I know*). Dans le cas dans la présente étude, les expressions figées devaient être utiles non seulement au plan didactique, mais également aux plans de la littératie et de la communication. En effet, il s'agissait de relever des expressions que les élèves étaient susceptibles de rencontrer dans leurs manuels scolaires et dans la littérature jeunesse qui leur est adressée, ainsi que dans la langue courante. Quant aux types d'expressions figées, ils étaient variés (les EF occupaient différentes fonctions grammaticales). Effectivement, à l'école et dans la vie de tous les jours, les enfants sont confrontés à des expressions de nature variées, et ils ont avantage à les reconnaître et à les employer pour mieux comprendre et mieux communiquer en français. Les expressions figées choisies devaient donc avoir également une valeur pragmatique et communicative.

La deuxième adaptation concernait la provenance des expressions figées. Alors que les EF du test de Schmitt ont été extraites de manuels scolaires destinés aux adultes et des études de deux chercheurs, les EF destinées à la conception de notre test ont été relevées dans des manuels scolaires pour enfants et de littérature jeunesse.

La présentation des items est aussi quelque peu différente dans notre instrument de mesure. Dans le test de Schmitt *et al.*, les expressions figées à compléter étaient suivies d'une paraphrase, mise entre parenthèses. Dans notre test, aucune paraphrase n'était fournie, notamment en raison des difficultés que cela aurait posées à la fois pour la création du test et pour les élèves. En effet, pour certaines expressions figées, il était presque impossible de trouver une paraphrase à la fois plausible et éclairante pour les élèves. Par exemple, l'expression « être à la hauteur » est définie comme suit dans *Le Petit Robert* : « faire preuve de compétence, d'efficacité ». Toutefois, cette paraphrase est assez vague et très littérale, alors que l'expression « être à la hauteur » est plutôt figurée ; de plus, les mots « compétence » et « efficacité » auraient probablement posé une difficulté supplémentaire aux plus jeunes élèves. Quant à l'expression « au fur et à mesure que », elle est, dans le même ouvrage, paraphrasée par la locution « à mesure que ». Or, bien que cette expression soit tout à fait synonyme, elle donne un indice trop évident de la réponse attendue à cause du mot « mesure » qui revient dans les deux expressions.

En somme, le choix de ne pas fournir d'équivalent a été fait, d'une part, parce qu'il était très difficile de trouver des paraphrases transparentes, concises et d'un niveau de langue équivalent à celui de la EF recherchée. D'autre part, cette méthode évitait d'alourdir et de compliquer la tâche par des paraphrases laborieuses et imprécises ou, au contraire, de faciliter la tâche par des indices trop évidents.

Autre différence à souligner : les expressions figées dans notre test n'ont pas été mises en caractères gras. Ce choix a été fait pour éviter d'attirer l'attention des élèves uniquement sur l'expression à compléter, alors que le contexte entourant les expressions est important. En effet, pour être capables de compléter les EF, ils devaient lire toutes les phrases de mise en contexte, de même que les phrases lacunaires.

Enfin, dans le test de Schmitt, les phrases lacunaires formaient un texte suivi (dans la première partie) et un dialogue (dans la deuxième partie). Cependant, dans notre test, les phrases étaient présentées séparément et non sous forme de texte,

pour éviter que les expressions laissées incomplètes n'entravent la compréhension et ne nuisent à la complétion des expressions subséquentes.

Les thèmes choisis pour la construction des items ont également été adaptés pour correspondre au vécu des enfants. Nous avons donc fait attention de créer des phrases susceptibles d'intéresser les élèves, en apportant aussi quelques références socioculturelles à la portée des enfants (fête de fin d'année, jeux olympiques, voyages, etc.).

4.7.3 Questionnaires complémentaires

Parallèlement au test, deux questionnaires complémentaires se sont avérés nécessaires pour mesurer l'influence de divers facteurs sur l'acquisition des EF. Le premier questionnaire consistait en une fiche de renseignements permettant de recueillir des renseignements sur le niveau scolaire et l'âge des élèves, la ou les langues parlées avec leurs parents, leurs frères et sœurs, leurs amis ainsi que l'année scolaire correspondant à leur intégration dans une classe régulière en français (la « classe régulière » étant définie par opposition à la « classe d'accueil », qui est vouée à la francisation des élèves). Ce questionnaire a été administré aux élèves au moment du test. Sur les 165 élèves interrogés, 142 d'entre eux (86 %) ont déclaré avoir fréquenté une classe régulière francophone depuis la maternelle. Les autres élèves, dans une proportion moindre, avaient commencé en 1^{re} année (11 élèves), en 2^e année (4 élèves), en 3^e année (2 élèves), en 4^e année (1 élève) et en 5^e année (2 élèves). Ces chiffres indiquent que pour la grande majorité des élèves (tous groupes linguistiques confondus), le français était la langue dans laquelle ils avaient été initiés aux bases de la littératie, c'est-à-dire la lecture et l'écriture (sauf peut-être quelques cas exceptionnels qui auraient commencé à apprendre à lire avec leurs parents à la maison avant d'entrer à l'école).

Le second questionnaire s'adressait aux enseignants titulaires. Il s'agissait d'un court sondage visant à savoir si des expressions figées avaient été enseignées au cours de l'année scolaire, et si oui, si cet enseignement avait été fait de façon formelle (avec exercices/évaluations) ou informelle (sans exercices/évaluations). Des exemples d'expressions étaient donnés en bas de page. Les enseignants ont

répondu au sondage lors de la passation du test par leurs élèves. Sur les sept enseignants interrogés, six ont affirmé avoir attiré l'attention des élèves sur certaines expressions de façon sporadique, notamment à travers les activités de lecture. Un seul enseignant a déclaré avoir dispensé un enseignement formel des expressions figées (par le biais d'exercices seulement). Donc, la plupart des participants n'avaient été soumis à aucun exercice ni aucun test portant sur les expressions figées depuis le début de l'année scolaire.

4.8 Validation du test

4.8.1 Échantillon

Préalablement à la collecte de données, le test a fait l'objet d'une validation dans une classe régulière de vingt-six (26) élèves de 6^e année. Tous les élèves, à une exception près, avaient le français comme langue première et n'éprouvaient pas de difficultés d'apprentissage. Il y avait 11 filles et 15 garçons. L'un des élèves avait un trouble envahissant du développement (TED); il était également le seul élève ayant deux langues premières, le français et l'anglais. Cela dit, nous reconnaissons qu'idéalement, il aurait été préférable de valider notre test auprès d'un échantillon se rapprochant encore davantage de la population étudiée (c'est-à-dire un groupe d'élèves comprenant un plus grand éventail de niveaux scolaires et de groupes linguistiques). Cependant, il était très difficile de trouver une école et des enseignants (autres que ceux déjà sollicités pour la collecte principale) disposés à se porter volontaires pour notre validation. Nous estimons que l'étape de la validation a tout de même été pertinente, car parmi les participants à cette recherche, les élèves francophones de 6^e année sont ceux qui, en principe, avaient atteint le niveau le plus élevé de compétences en français, par rapport aux élèves allophones et francophones des niveaux scolaires inférieurs. L'échantillon étudié a donc été utile dans la mesure où il a permis de cerner les difficultés susceptibles de survenir lors de l'expérimentation et d'anticiper les résultats possibles, du moins chez les élèves les plus avancés.

4.8.2 Milieu scolaire

L'école où a eu lieu la validation du test était une école publique située en banlieue de Montréal, dans un quartier défavorisé. La validation a été effectuée en classe, en présence de l'expérimentatrice et de l'enseignante titulaire. À la fin du test, les élèves ont répondu par écrit à un court sondage visant à cerner leur perception du niveau de difficulté du test. La grande majorité des élèves (22 sur 26) ont jugé le test comme étant « plutôt facile » (17 élèves) ou « très facile » (5 élèves), et les résultats obtenus reflétaient bien leur impression : en moyenne, les élèves ont réussi 14,85 items sur 18 (les réponses étaient jugées comme « réussies » lorsque l'expression recherchée était phonétiquement correcte). Le barème de correction sera expliqué plus loin.

Compte tenu des questions et des commentaires des élèves lors de la validation du test, les questions figurant sur la fiche de renseignements ont été simplifiées en vue de la collecte. Quant aux items du test, ils sont demeurés inchangés.

4.9 Collecte de données

Avant l'administration du test aux élèves, chaque assistant de recherche a reçu un protocole de passation détaillé indiquant la marche à suivre pour la passation du test. Au moment du test, l'expérimentateur a distribué un exemplaire du test à chaque élève. Il a demandé aux enfants, dans un premier temps, de remplir la première page du document, qui correspondait à la fiche de renseignements sur l'élève (année scolaire, date de naissance, langues parlées avec les parents, avec les frères et sœurs, avec les amis, année du début de scolarisation régulière en français). Dans un deuxième temps, l'expérimentateur a lu la consigne du test aux élèves. Cette dernière se lisait comme suit : « Lis bien les phrases suivantes et complète les mots incomplets sur les tirets ». Pour aider les élèves à comprendre le genre de réponses attendues, il leur a donné deux exemples (qui étaient écrits sur les copies des élèves), s'est assuré que tous les élèves avaient compris la consigne et a répondu à leurs questions, le cas échéant.

Par la suite, les élèves devaient lire eux-mêmes les items du test en silence et les compléter avec les mots manquants. Le test a été effectué de façon individuelle. Les élèves n'avaient accès à aucun ouvrage de référence.

L'expérimentateur était tenu de ne pas répondre aux questions des élèves concernant les items en tant que tels (par exemple, les questions sur le sens des mots des items, sur l'orthographe, etc.).

4.10 Compilation des résultats

Puisque le principal objectif de notre recherche était de déterminer les connaissances des élèves en matière d'expressions figées, il nous a fallu établir un barème de correction reflétant divers niveaux de maîtrise des EF. En effet, les réponses des élèves pouvaient refléter une connaissance minimale (reconnaissance), une connaissance pragmatique (contextuelle), ou une connaissance orthographique. Les réponses ont été compilées dans une base de données Excel, puis codées selon le barème suivant :

0 : aucune réponse

1 : connaissances insuffisantes de l'expression

- a) mot(s) français ne faisant pas partie de l'expression
- b) mot(s) inexistant(s) en français
- c) réponse partielle (1 mot correct sur 2, dans le cas des expressions comprenant deux mots à compléter)
- d) réponse illisible
- e) réponse incohérente ou ambiguë
- f) anomalies (par exemple : l'élève a changé des lettres ou des mots déjà fournis sur sa copie, ou il a mis un ou plusieurs mots en trop, etc.)

2 : seuil de connaissance minimale de l'expression

L'élève a vraisemblablement reconnu l'expression recherchée, mais l'a écrite de façon approximative, avec au moins une erreur de transcription phonétique. Toutefois, l'expression est reconnaissable en contexte malgré la ou les erreurs de

phonétique. Puisqu'il s'agit d'un test écrit, l'erreur phonétique se traduit par une faute d'orthographe ou de grammaire.

3 : seuil de connaissance pragmatique de l'expression

L'élève connaît l'expression et est capable de la produire correctement en contexte. Phonétiquement, l'expression produite correspond tout à fait à l'expression recherchée. Toutefois, elle comporte au moins une faute d'orthographe ou de grammaire.

4 : seuil de connaissance orthographique et grammaticale de l'expression

L'élève connaît l'expression recherchée et est capable de la produire correctement en contexte. Phonétiquement, l'expression produite correspond tout à fait à l'expression recherchée. De plus, il a su l'écrire sans faute d'orthographe ou de grammaire.

Ce barème a été établi en fonction des spécificités de notre étude, de façon à ce que chaque réponse reflète un « niveau » de connaissance. Toutefois, certaines réponses fournies par les élèves auraient pu être classées dans plus d'une catégorie. Afin de minimiser la subjectivité, nous avons fait appel à quatre informateurs afin de conclure une entente interjuge. Les quatre informateurs étaient renseignés sur les objectifs de l'étude, avaient accès au test original et au corrigé, et possédaient des connaissances linguistiques suffisantes pour donner un jugement éclairé. Pour chacun des items, il y avait accord lorsque 3 évaluateurs sur 5 proposaient le même code de correction. Sur les 45 types de réponses soumises à l'évaluation interjuge, 38 ont pu être codées ; les 7 autres types de réponses ont été classés comme des items « sans correction » (code 9).

Pour répondre aux deux questions de recherche, le barème original a été simplifié. Dans le nouveau barème, le code « 1 » regroupe les codes 0, 1, et 2 du barème original et représente les réponses jugées insuffisantes. Le code « 2 », quant à lui, regroupe les codes 3 et 4 du barème original et correspond aux réponses jugées satisfaisantes.

- 1 : mauvaise réponse (codes 0, 1, et 2 du barème original)
- 2 : bonne réponse (codes 3 et 4 du barème original)
- 3 : réponse non évaluée (items ambigus qui n'ont pas pu être notés même après l'entente interjuge, faute de consensus)

Rappelons que la présente étude vise à mesurer les connaissances des élèves en matière d'expressions figées. Même s'il existe plusieurs niveaux de connaissance, nous considérons que le seuil le plus significatif est le seuil pragmatique compte tenu de la population visée (enfants de niveau primaire, forte proportion d'allophones). En effet, ce seuil est d'une importance capitale, car il correspond à une connaissance fonctionnelle des expressions figées : il indique la capacité des élèves à reconnaître, à comprendre et à employer correctement les expressions figées en contexte. C'est pourquoi nous avons jugé plus pertinent de fixer le seuil de réussite en fonction des connaissances pragmatiques, plutôt que des connaissances minimales (qui sont souvent trop fragiles ou approximatives) ou que des connaissances orthographiques et grammaticales (qui démontrent une plus grande maîtrise du code écrit, mais pas forcément une connaissance plus poussée des expressions figées).

Il est à noter que Schmitt *et al.* n'ont pas évalué la performance de leurs participants en fonction d'un barème de correction précis. Étant donné la nature de leur étude, un tel barème n'était pas nécessaire. Toutefois, dans le cas de la présente étude, le barème était nécessaire étant donné l'objet de l'étude ainsi que les particularités phonétiques, grammaticales et orthographiques du français.

4.11 Le calcul du score

Dans le cadre de notre analyse, le score est défini comme étant la proportion d'items réussis sur le nombre total d'items ayant été évalués. Afin de mesurer la performance individuelle des élèves au test, il s'agira donc de calculer le nombre de bonnes réponses (codées « 2 ») obtenues sur un total de 18 (puisque le test comprenait 18 items), et de convertir ce nombre en pourcentage.

4.12 Analyses

Des analyses de variance (ANOVA) nous permettront de vérifier s'il existe un lien entre la performance des élèves et les facteurs individuels suivants : l'année scolaire, le groupe linguistique et le sexe. Nous examinerons aussi l'impact possible de la fréquence des expressions figées, des mots clés (mots tronqués) et du nombre de concurrents des mots clés.

La fréquence des mots clés ainsi que le nombre de concurrents plus fréquents seront vérifiés dans Manulex, une base de données lexicale formée à partir de manuels scolaires français (élèves de 6 à 11 ans)⁶. Quant à la fréquence des expressions, elle sera mesurée dans Lexique⁷ (car il n'est pas possible de la vérifier dans Manulex). Lexique est une base de données comprenant un corpus de français écrit (auteurs classiques) et un corpus oral (films).

4.13 Conclusion du chapitre

Ce chapitre avait pour objectif de présenter la démarche méthodologique et d'en expliquer les fondements en s'appuyant sur la recherche de Schmitt, Dörnyei, Adolphs et Durow (2004). Chaque étape de la démarche a donc fait l'objet d'une description détaillée, en mettant l'accent sur les principales similitudes et différences entre la présente recherche et celle de Schmitt *et al.* Les questions de recherche, l'âge des participants et la langue cible, qui constituent les différences fondamentales entre les deux recherches, ont donné lieu à d'importantes adaptations, en particulier sur le plan de la conception de l'instrument de mesure. Par conséquent, le type d'expressions figées, la présentation des items et le barème d'évaluation ont été modifiés en fonction des particularités de la présente recherche.

⁶ Version française téléchargeable à l'adresse suivante : <http://unpc.univ-lyon2.fr/~lete/manulex/index.html>

⁷ Base de données accessible sur <http://www.lexique.org/>

Le prochain chapitre sera consacré à la présentation et à l'analyse des données recueillies.

CHAPITRE V

ANALYSE STATISTIQUE ET INTERPRÉTATION DES

RÉSULTATS

L'objectif du présent chapitre est de présenter et d'interpréter les résultats de l'analyse statistique effectuée à partir de l'information recueillie lors de la collecte de données. Dans un premier temps, nous répondrons à nos questions de recherche qui concernent l'influence de l'année scolaire, du groupe linguistique et du sexe sur la performance des élèves. Nous allons également vérifier l'influence possible de trois facteurs linguistiques sur la réussite des élèves, à savoir la fréquence des expressions, la fréquence des mots ciblés et le nombre de concurrents des mots ciblés. Dans un deuxième temps, nous ferons une analyse par item. Celle-ci nous permettra, d'une part, de déterminer quels ont été les items les mieux réussis et les moins bien réussis et d'autre part, de savoir si les taux de réussite des items reflètent des écarts de performance entre les années scolaires et entre les groupes linguistiques.

5.1 Analyse par année scolaire

Rappelons que la première question de recherche visait à déterminer s'il y avait une corrélation positive entre le niveau scolaire et le rendement au test. La première étape de l'analyse statistique consistait donc à comparer les taux de réussite des élèves de 3^e, 4^e, 5^e et 6^e année.

Le tableau ci-dessous montre la performance moyenne de l'ensemble des élèves pour chaque année scolaire. La colonne « N » indique le nombre d'élèves par année scolaire, la « moyenne » correspond à la proportion d'items réussis (exprimée en pourcentage), et l'abréviation É.T. fait référence à l'écart type.

Tableau 5.1 Taux de succès par année scolaire

<i>Année</i>	<i>N</i>	<i>M</i>	<i>É-T</i>
3	27	31,89	15,68
4	33	45,45	15,62
5	45	57,90	22,89
6	46	66,43	18,59
DGA	14	35,32	20,51

Un coup d'oeil au tableau suffit pour constater une nette progression de la proportion d'items réussis d'une année scolaire à l'autre, de la 3^e à la 6^e année. En ce qui concerne les élèves en difficulté d'apprentissage, leur performance moyenne (36 % d'items réussis) se rapproche de celle des élèves de 3^e année. Un test ANOVA a permis de confirmer que l'effet de l'année scolaire était positivement significatif ($F(3, 147) = 21.54, p < 0.0001$)¹.

¹ Le groupe DGA n'a pas été inclus dans cette analyse parce que cette classe ne correspond pas à un niveau scolaire en particulier.

Les performances des élèves ont ensuite été soumises à une comparaison deux à deux avec ajustement de Tukey-Kramer. Ce test a révélé une progression significative entre les élèves de 3^e et de 4^e année, et entre les 4^e année et le dernier cycle du primaire, c'est-à-dire la 5^e et la 6^e année regroupées. Bien que les élèves de 6^e année aient réussi, en moyenne, une plus grande proportion d'items que les élèves de 5^e année, la différence de performance de ces deux groupes n'était pas statistiquement significative. Les valeurs statistiques sont les suivantes :

$$3^{\text{ième}} < 4^{\text{ième}} < (5^{\text{ième}} \text{ et } 6^{\text{ième}});$$

$$3 < 4 \text{ (} p = 0.0272 \text{)}$$

$$3 < 5 \text{ (} p < 0.0001 \text{) et } 6 \text{ (} p < 0.0001 \text{)}$$

$$4 < 5 \text{ (} p = 0.0287 \text{) et } 6 \text{ (} p < 0.0001 \text{)}$$

L'influence significative du facteur « année scolaire » semble confirmer ce qui a déjà établi concernant l'évolution intellectuelle des enfants au fil des années. En effet, rappelons que les enfants d'âge scolaire voient leurs aptitudes cognitives progresser rapidement (Blaye et Lemaire 2007; Nippold 1998). L'enfance est également une période où les enfants développent leurs connaissances pragmatiques, linguistiques et extralinguistiques, qui sont essentielles à l'acquisition des expressions (Iralde 2004). De plus, la progression des enfants dans leur scolarisation s'accompagne d'une évolution importante au plan des connaissances lexicales vu leur contact prolongé avec la langue écrite, et ce, surtout pendant la période allant de l'âge de 8 à 11 ans (Lété 2004). En effet, d'une année scolaire à l'autre, le vocabulaire des enfants non seulement s'accroît en nombre, mais il devient de plus en plus solidement ancré dans leur mémoire. Rappelons à cet effet qu'il est essentiel qu'un mot soit suffisamment consolidé dans la mémoire du locuteur pour que celui-ci soit capable de bien le comprendre et de bien l'employer (Lété 2004). Puisque le test consistait à produire les expressions à l'écrit (vocabulaire actif), il est normal d'observer un écart significatif entre la performance des élèves plus vieux et celle des élèves plus jeunes. Donc, les résultats de notre expérimentation peuvent signifier que de la 3^e à la 6^e année, beaucoup d'expressions passent du vocabulaire

passif (compris) au vocabulaire actif (compris et assez bien mémorisé pour être employé). Cela dit, mentionnons que nos résultats corroborent également la théorie de Wray (2002) voulant que pendant la période allant de l'âge de 8 à 17 ans environ, les jeunes locuteurs natifs emploieraient une proportion de plus en plus importante de langage figé : il est probable que cette évolution soit le reflet d'une progression dans leur niveau de maîtrise des expressions figées. Toutefois, selon Wray (2002), cette évolution serait propre aux locuteurs natifs. Alors comment expliquer que cette tendance se manifeste de façon aussi marquée dans un groupe comprenant une proportion de 28% d'enfants allophones? Il est possible que des locuteurs non natifs ayant appris leur langue seconde à l'âge préscolaire (comme la plupart des élèves allophones de notre expérimentation), suivent une évolution à peu près semblable à celle de leurs pairs francophones en ce qui concerne la capacité de comprendre et d'employer des expressions figées. Nous verrons justement, dans la prochaine sous-section, si les groupes linguistiques se sont distingués au plan de la performance au test.

5.2 Analyse par groupe linguistique

La deuxième question de recherche visait à savoir si le groupe linguistique avait influencé la réussite au test. Le tableau suivant montre, pour chaque niveau scolaire, le score moyen des élèves pour chaque groupe linguistique.

Tableau 5.2 Taux de réussite par année et par groupe linguistique

Année	<i>Francophones</i>			<i>Allophones</i>		
	N	M	É-T	N	M	É-T
3	16	31,60	16,45	11	32,32	15,28
4	19	44,15	17,42	14	47,22	13,21
5	25	54,70	23,61	18	62,28	21,72
6	33	66,50	19,91	13	66,24	15,26

Contrairement aux attentes exprimées lors de la formulation des questions de recherche, les scores des élèves se sont révélés très semblables d'un groupe linguistique à l'autre, et ce, peu importe le niveau scolaire. Un test ANOVA a effectivement confirmé qu'il n'y avait pas de corrélation entre le facteur « groupe linguistique » et la performance des élèves. L'interaction entre ces deux facteurs a également été mesurée. Voici les calculs effectués :

Comparaison des moyennes du score selon l'année, le groupe linguistique et leur interaction

Année : $F(3, 150) = 19,44$, $p < 0.0001$

Groupe linguistique : $F(1, 150) = 0.71$, $p = 0.400$

Interaction : $F(3, 150) = 0.33$, $p = 0.803$

Plusieurs raisons peuvent expliquer l'absence de différence significative entre les groupes linguistiques. D'abord, rappelons que la grande majorité des élèves allophones étaient inscrits dans une classe régulière francophone depuis la maternelle. C'est donc dire que même si ces élèves n'ont pas le français comme première langue, ils ont, en revanche, fait la totalité de leur scolarisation en français et ont appris à lire et à écrire en français en même temps que leurs pairs francophones. Ils reçoivent donc une quantité importante d'input oral et écrit depuis l'âge de 5 ans, et évoluent dans un contexte authentique d'apprentissage, ce qui favorise l'apprentissage et l'emploi d'EF (Wray 2002). Par conséquent, il est possible que les enfants allophones (de la 3^e à la 6^e année) aient acquis des compétences suffisantes en français pour réussir aussi bien que leurs collègues francophones au test sur les EF. De plus, il s'agissait d'un test écrit qui ciblait des EF caractéristiques de la langue écrite, alors que c'est surtout le contact avec le français oral à la maison qui diffère entre ces deux groupes. En effet, rappelons que les groupes linguistiques avaient été déterminés en fonction de la langue parlée le plus souvent avec les parents. Bien que cette décision soit justifiée, force est d'admettre que même si un enfant parle une autre langue que le français dans son milieu familial, il peut néanmoins développer de fortes compétences en français, surtout s'il évolue

dans un milieu scolaire francophone et qu'il maintient un contact avec le français à l'extérieur de l'école (par exemple, en parlant avec ses amis ou en lisant en français). Une autre explication possible est que l'exposition à la langue écrite (que nous présumons équivalente chez les deux groupes linguistiques) soit plus importante que l'exposition à la langue orale pour l'acquisition des expressions figées. Si l'on avait étudié une population composée de deux groupes, francophones et d'allophones, dont la principale différence était la quantité d'input écrit en français, peut-être aurions-nous obtenu un résultat différent. Cela dit, mentionnons que la vaste majorité de ces enfants étaient bilingues ou plurilingues. L'opposition entre les francophones et allophones n'était donc pas aussi tranchée que si nous avions affaire à un groupe d'enfants unilingues francophones et à un groupe d'allophones bilingues. Le statut linguistique des élèves étant ainsi nuancé, les écarts de performances entre les groupes linguistiques ont été atténués.

5.3 Analyse par sexe

Nous avons également vérifié si le sexe avait eu une influence quelconque sur la performance au test. Dans la commission scolaire où nous avons effectué notre enquête, ainsi que dans l'ensemble du territoire québécois (référence aux résultats scolaires par sexe sur la page Web du MELS), les garçons obtiennent généralement des résultats scolaires plus faibles que ceux des filles. Nous avons donc tenu compte de ce facteur dans nos analyses. Le tableau suivant montre la performance des filles et des garçons.

Tableau 5.3 Taux de réussite par année et par sexe

Année	<i>Filles</i>			<i>Garçons</i>		
	N	M	É-T	N	M	É-T
3	16	32,29	16,88	11	31,31	14,55
4	17	44,12	15,27	16	46,88	16,35
5	25	62,61	20,70	19	51,46	24,70
6	28	67,26	20,25	13	65,12	16,15

Les moyennes ont été comparées selon l'année, le sexe et leur interaction :

Année : $F(3, 141) = 19.89, p < 0.0001$;
 Sexe : $F(2, 141) = 0.97, p = 0.3263$;
 Interaction : $F(6, 141) = 1.05, p = 0.3736$;

Or, il s'est avéré que la différence entre les résultats des garçons et des filles n'était pas significative. Par conséquent, nous ne retiendrons pas ce facteur dans les analyses par item qui suivront.

5.4 Analyse de la fréquence des expressions, de la fréquence des mots clés et du nombre de concurrents

Rappelons que la fréquence des mots ciblés (tronqués) a été vérifiée dans Manulex, une base de données lexicale formée à partir de manuels scolaires français (élèves de 6 à 11 ans)². En effet, il est possible que certains élèves puissent avoir « deviné » des mots tronqués, surtout dans le cas des mots fréquents ou des mots ayant peu ou pas de concurrents. Quant à la fréquence des expressions, il n'était pas possible de la vérifier dans Manulex ; nous l'avons donc mesurée dans Lexique³, une base de données comprenant un corpus de français écrit (auteurs classiques) et un corpus oral (films). Il est à noter que toutes les mesures de fréquence auxquelles nous ferons référence dans notre discussion (fréquence des expressions et des mots) correspondent au nombre d'occurrences par million. Nous avons aussi mesuré l'influence du nombre de concurrents, c'est-à-dire le nombre de mots plus fréquents commençant par la ou les mêmes lettres qu'un mot donné. Par exemple, le mot ciblé commençant par la syllabe qua_____ (« quatre », dans l'expression « les quatre coins du monde ») n'avait qu'un seul concurrent plus

² Version française téléchargeable à l'adresse suivante : <http://unpc.univ-lyon2.fr/~lete/manulex/index.html>

³ Base de données accessible sur <http://www.lexique.org/>

fréquent dans Manulex (« quand »). Le Tableau 5.4 (page suivante) donne un aperçu des caractéristiques linguistiques de chaque item du test.

Tableau 5.4 Caractéristiques des items : fréquence des EF, fréquence des mots ciblés et nombre de concurrents des mots ciblés.

Item	Expression	Frq. Lexique livres	Frq. Lexique films	Frq. mot clé le moins frq ⁴	Nombre de concurrents plus frq ⁵
1	à l' <u>occasion</u> de	99	14	40,53	0
2	tout le monde	2007	3501	595,98	1
3	<u>plus</u> ou <u>moins</u>	595	97	317,90	17
4	en ce qui <u>concerne</u>	114	84	7,71	80
5	être à la <u>hauteur</u>	13	30	54,86	2
6	bel et <u>bien</u>	116	48	2755,97	0
7	tout au <u>long</u>	180	32	143,43	25
8	être aux <u>prises</u> avec qqch	69	6	1,04	44
9	de <u>manière</u> à ce que	4	1	135,19	14
10	au fur et à <u>mesure</u>	31	211	41,99	108
11	à tour de <u>rôle</u>	90	16	52,55	0
12	à toute <u>vitesse</u>	148	40	149,80	10
13	depuis belle <u>lurette</u>	17	8	0	138
14	à portée de la <u>main</u>	50	3	15,76	14
15	les quatre coins du <u>monde</u>	13	7	442,98	10
16	sain et <u>sauf</u>	10	69	2,47	75
17	<u>flambant</u> neuf	16	15	1,01	55
18	à <u>perte</u> de <u>vue</u>	110	25	14,47	66

⁴ Ce chiffre a été arrondi dans la discussion afin d'alléger le texte.

⁵ Nombre de concurrents du mot clé de l'EF ayant le plus de concurrents.

Dans un premier temps, nous avons mesuré l'effet de la fréquence des expressions dans la base de données Lexique. Nous avons d'abord consulté la base de données regroupant le lexique provenant de livres, afin de mesurer la fréquence des expressions dans la langue écrite. Nous avons ensuite consulté la base de données formée à partir des dialogues de films pour estimer la fréquence des expressions dans la langue parlée. Cependant, les tests statistiques que nous avons réalisés n'ont pas révélé de corrélation significative entre la performance des participants et la fréquence des expressions. Ce résultat ne corrobore donc pas les recherches de Laval (2003), Nippold et Rudzinski (1993) et Levorato et Cacciari (1992), où la fréquence avait joué un rôle significatif dans la connaissance des expressions chez les enfants. Signalons toutefois qu'il s'agit d'une base de données lexicale reflétant la langue parlée et écrite chez les adultes. Par conséquent, les données que nous y avons recueillies ne représentent pas réellement le vocabulaire auquel les enfants sont exposés. Il est donc possible que la fréquence des expressions ait eu un effet sur la réussite des élèves ; or les données statistiques dont nous disposons ne nous permettent pas de confirmer cette hypothèse.

Dans un deuxième temps, nous avons mesuré l'influence possible de la fréquence des mots ciblés. Nous avons examiné le lien possible entre la réussite des élèves des différents niveaux scolaires et la fréquence des mots ciblés (dans le cas des items ayant deux mots tronqués, c'est le mot le moins fréquent qui a servi à l'analyse). Les résultats n'ont pas montré de différence significative. Nous avons donc tenté une seconde analyse, en regroupant cette fois-ci les mots-clés par catégories :

- 1) mots rares (0 à 10 occurrences par million)
- 2) mots fréquents (11 à 99 occurrences par million)
- 3) mots très fréquents (100 et +)

L'analyse a révélé une influence significative de la fréquence des mots ciblés : les items contenant des mots plus fréquents ont été mieux réussis que les items contenant des mots rares chez les élèves de 4^e année ($p=0,021$), de 5^e année

($p=0,006$) et de 6^e année ($p=0,016$). Nous avons ensuite vérifié si ce facteur avait exercé une influence dans les sous-groupes. Les items contenant des mots fréquents ont été mieux réussis que les items comprenant des mots rares, et ce, chez les 4^e année allophones ($p=0,005$), les 5^e année francophones ($p=0,024$) et allophones ($p=0,002$), ainsi que les 6^e année francophones ($p=0,019$) et allophones ($p=0,012$).

Dans un troisième temps, nous avons mesuré l'influence du nombre de concurrents par année et par groupe linguistique. Les calculs effectués (test ANOVA) montrent que le nombre de concurrents a eu un effet significatif chez les 3^e année francophones ($p=0,026$), chez les 4^e année francophones et allophones ($p=0,006$) ainsi que chez les 5^e année francophones ($p=0,031$). Nous avons également procédé à une analyse par catégories :

- 1) peu de concurrents (0 à 2)
- 2) quelques concurrents (3 à 17)
- 3) beaucoup de concurrents (20 et plus).

Les résultats de notre analyse montrent que les items ayant peu de concurrents ont été moins bien réussis que les items ayant beaucoup de concurrents chez les 3^e année francophones ($p=0,002$) et allophones ($p=0,032$), chez les 4^e année francophones ($p < 0,001$) et allophones ($p=0,005$), chez les 5^e année francophones ($p=0,01$), ainsi que chez les 6^e année francophones ($p=0,026$) et allophones ($p=0,014$). Les items ayant peu de concurrents ont également été mieux réussis que les items ayant quelques concurrents chez les 3^e année francophones ($p=0,027$) de même que chez les 4^e année francophones ($p=0,018$).

Les résultats de nos analyses montrent que le nombre de concurrents ainsi que la fréquence des mots clés ont eu un effet important sur la réussite des élèves au test. Devant ce constat, il y a lieu de s'interroger sur la signification de ce phénomène. L'une des explications possibles est que les élèves ont probablement plus de facilité à remarquer, comprendre et produire les expressions qui contiennent des mots familiers que les expressions qui contiennent des mots rares. Cela dit, si les élèves performant mieux aux items dont les mots ciblés sont fréquents et dont le

nombre de concurrents est minime, on se doit d'envisager la possibilité que certains élèves puissent avoir réussi à compléter certaines expressions sans les connaître vraiment, c'est-à-dire en écrivant le ou les mots qui leur semblaient les plus plausibles sémantiquement et grammaticalement. Les données dont nous disposons ne nous permettent pas d'expliquer avec certitude les raisons pour lesquelles la fréquence et le nombre de concurrents des mots cibles sont autant liés à la réussite des items ; il serait par contre intéressant de se pencher sur cette question lors de recherches ultérieures.

5.5 Analyse par item

5.5.1 Performance des élèves aux items

Dans cette section, nous avons classé les items selon les taux de réussite : les items réussis par une forte majorité d'élèves (dont les taux de succès avoisinent les 90 %), les items réussis par une majorité d'élèves (c'est-à-dire réussis par plus de la moitié des participants), les items réussis par une minorité d'élèves (réussis par moins de la moitié des élèves), ainsi que les items réussis par une faible minorité d'élèves (taux de réussite de 20 % et moins). Nous interpréterons les taux de réussite généraux des items en faisant référence à la fréquence des mots ciblés et au nombre de concurrents, puisque ces deux facteurs ont joué un rôle significatif sur la performance des élèves. Avant de poursuivre la discussion, nous présenterons ici le Tableau 5.5, qui regroupe les taux de réussite des élèves par année scolaire.

Tableau 5.5 Taux de réussite de chaque item selon l'année scolaire

Item	Expression	Taux global	Rang	3 ^e	4 ^e	5 ^e	6 ^e	DGA
1	à l'occasion de	58,18	8	37,0	54,5	64,4	69,6	50
2	tout le monde	91,46	2	85,2	90,9	95,6	95,7	71,4
3	plus ou moins	62,80	6	25,9	57,6	80,0	78,3	35,7
4	en ce qui concerne	49,70	9	37,0	27,3	53,3	76,1	28,6
5	être à la hauteur	88,48	3	77,8	90,9	86,7	100	71,4
6	bel et bien	61,82	7	22,2	57,6	64,4	89,1	50,0
7	tout au long	58,18	8	33,33	48,5	66,7	80,4	28,6
8	être aux prises avec qqch	4,08	17	0,0	6,1	2,2	6,5	0,0
9	de manière à ce que	42,59	11	14,8	27,3	48,9	63,0	35,7
10	au fur et à mesure	65,45	5	29,6	57,6	80,0	84,8	42,9
11	à tour de rôle	92,73	1	88,9	97,0	86,7	97,8	92,9
12	à toute vitesse	68,10	4	66,7	63,6	71,1	78,3	28,6
13	depuis belle lurette	12,73	16	3,7	0,0	15,6	28,3	0,0
14	à portée de la main	46,67	10	11,1	42,4	60,0	63,0	28,6
15	les quatre coins du monde	38,18	12	18,5	36,4	51,11	43,5	21,4
16	sain et sauf	34,55	14	18,5	24,2	46,7	41,3	28,6
17	flambant neuf	20,73	15	3,7	15,2	20,0	37,0	14,3
18	à perte de vue	35,76	13	0,0	21,2	48,9	63,0	7,1

A) Items réussis par une forte majorité d'élèves

Les items 11, 2, 5 ont été réussis par presque tous les élèves. Les taux de réussite oscillent entre 88,48 % et 92,73 %.

11 : à tour de rôle

2 : tout le monde

5 : être à la hauteur

L'item 11 (à tour de rôle) a été le mieux réussi de tous avec un taux de réussite de 92,73 %. L'absence de concurrents peut expliquer le succès des enfants à compléter cette expression : le mot ciblé (rôle), qui commençait par la syllabe « rô », n'avait aucun concurrent plus fréquent dans Manulex, ce qui signifie qu'en français les lettres « rô », sont le plus souvent suivies par les lettres « le ». Donc, il est fort probable que ce mot soit le premier (et peut-être le seul) auquel les enfants aient pensé. Il s'agit également d'un mot fréquent (53 occurrences par million). L'item 2 (tout le monde) a également été réussi par une très forte proportion d'élèves 91,46 %). « Monde » est un mot d'utilisation très fréquente (596 occurrences par million); de plus, la séquence « mon » n'avait qu'un seul concurrent plus fréquent dans Manulex. Les élèves ont également bien performé à l'item 5 (être à la hauteur) : 88,48 %. Ce mot assez fréquent dans les manuels scolaires (55 occurrences par million) et n'a que 2 concurrents plus fréquents dans Manulex. Donc, nous ne pouvons pas écarter la possibilité que certains enfants aient trouvé la bonne réponse sans connaître l'expression.

B) Items réussis par une majorité d'élèves

Les items 12, 3, 6 et 7 ont été réussis par plus de la moitié des élèves. Les taux de réussite varient entre 58,18 % (items 1 et 7) et 68,10 % (item 12).

12 : à toute vitesse

10 : au fur et à mesure

3 : plus ou moins

6 : bel et bien

7 : tout au long

1 : à l'occasion de

L'item 12 (à toute vitesse) a été complété avec succès par 68,10 % des participants. Les deux mots clés sont des mots très fréquents dans les manuels scolaires (ils ont respectivement 673 et 150 occurrences par million dans Manulex). Le mot « toute » n'a que 4 concurrents, alors que le mot « vitesse » en compte 10. L'item 10 (au fur et à mesure) a également été réussi par une forte proportion d'élèves, soit 65,45 %. Puisque le mot « fur » n'existe que dans l'expression « au fur et mesure », il était facile de reconnaître et de compléter cette expression. Quant à l'item 3 (plus ou moins), 62,80 % des élèves l'ont réussi. Les deux mots tronqués sont très fréquents (le mot « plus » apparaît 3723 fois et le mot « moins », 318 fois par million) et comptent un nombre plutôt restreint de concurrents (2 et 17, respectivement). Pourtant, près de 40 % des élèves n'ont pas réussi à compléter cette expression. La principale difficulté de cet item réside dans le petit nombre de lettres fournies, soit la première lettre de chaque mot porteur de sens : p_____ (« plus ») ou m_____ (« moins »). L'analyse du contexte était donc essentielle pour trouver l'expression appropriée. L'item 6 (bel et bien) a été réussi par 61,82 % des participants. Le mot « bien » est un mot extrêmement fréquent dans les manuels scolaires (2756 occurrences par million). Par ailleurs, il faut signaler que le mot ciblé (« bien ») ne comptait aucun concurrent plus fréquent dans Manulex : il est donc possible que certains élèves aient trouvé le bon mot « par hasard ». Les items 1 (à l'occasion de) et 7 (tout au long) ont connu le même taux de succès, soit 58,18 %. En ce qui a trait au premier item, le mot « occasion » est très courant (41 occurrences par million). Précisons également que l'amorce « occa_____ » n'a aucun concurrent plus fréquent dans Manulex, ce qui pourrait éventuellement avoir favorisé la chance chez certains élèves. Quant au second item, le mot clé (« long ») s'avère très fréquent (143 occurrences p. million). En revanche, il a un nombre élevé de concurrents (25), ce qui explique pourquoi une proportion assez importante d'élèves (42 %) n'a pas complété l'expression correctement.

C) Items réussis par une minorité d'élèves

Les items 16, 18, 15, 14 et 4 ont été réussis par moins de la moitié des élèves : les taux de réussite varient entre 34,55 % (item 16) et 49,70 % (item 4).

- 16 : sain et sauf
- 18 : à perte de vue
- 15 : les quatre coins du monde
- 9 : de manière à ce que
- 14 : à portée de la main
- 4 : en ce qui concerne

L'item 16 (sain et sauf) a connu un taux de succès de 34,55 % des élèves l'ont complété correctement. Le faible taux de succès peut être attribuable au fait que le mot « sauf » est peu fréquent (2 occurrences par million) et qu'il a un nombre très élevé de concurrents (75). Cela dit, il est possible que certains enfants n'aient pas « reconnu » l'expression à partir du mot « sain », d'autant plus que ce mot comporte de nombreux homonymes (saint, ceint, sein, etc.), ce qui peut prêter à confusion. L'item 18 (à perte de vue) a été réussi par 35,76 % des élèves. Les deux mots ciblés avaient un nombre de concurrents assez élevé dans Manulex (15 pour « perte » et 66 pour « vue »). Il est donc improbable que des élèves puissent avoir deviné les mots de l'expression. L'item 15 (les quatre coins du monde) a également été réussi par une minorité d'élèves (35,76 %). Pourtant, les deux mots ciblés étaient très fréquents dans la base Manulex (443 et 565 occurrences respectivement) et avaient peu de concurrents (1 et 10 respectivement). On peut attribuer le faible taux de réussite au fait qu'il y avait deux (2) mots tronqués à compléter. De plus, en ce qui a trait au second mot cible (« monde »), seule la première lettre était fournie, ce qui augmentait la difficulté de reconnaître le mot. En ce qui concerne l'item 9 (de manière à), son taux de succès est relativement faible (42,59 %) en dépit du fait que le mot clé (« manière ») soit très fréquent dans les manuels scolaires (136 occurrences) et que le nombre de concurrents ne soit pas très élevé (14). L'insuccès de cet item peut être attribuable au fait que la syllabe « ma » ait porté à confusion chez quelques élèves, qui auraient peut-être interprété « ma » comme un adjectif possessif. Enfin, la relative complexité grammaticale de l'expression (locution

conjonctive) pourrait être à l'origine de difficultés de compréhension de l'item. Quant à l'item 14 (à portée de la main), 46,67 % des participants ont trouvé la bonne réponse. Le premier mot cible (« portée ») est plutôt fréquent (16 occurrences dans Manulex) et n'a que 8 concurrents ; quant au second mot cible (« main »), il s'avère très fréquent (454 occurrences) et compte 14 concurrents. L'item 4 (en ce qui concerne) a été réussi par près de la moitié des élèves, soit 49,70 %. Le mot à compléter (« concerne ») est assez rare (8 occurrences p. million) et a un très grand nombre de concurrents (80) dans Manulex.

D) Items réussis par une faible minorité d'élèves

Les items les moins bien réussis sont les items 8, 13, et 17, avec des taux de réussite de 20,73 % et moins.

8 : être aux prises avec qqch

13 : depuis belle lurette

17 : flambant neuf

L'item 8 (être aux prises avec qqch) a vraisemblablement été trop difficile pour tous les élèves. Alors que le mot « avec » est très fréquent (3768) et n'a aucun concurrent, le mot « prises » n'a qu'une seule occurrence par million dans Manulex; de surcroît, ce mot a un nombre élevé de concurrents (44). Étant donné le nombre élevé de concurrents pour le mot « prises » (44), les chances que les élèves « devinent » le mot correctement étaient pratiquement nulles. Nous sommes donc assurée que les élèves ayant complété correctement les deux éléments de réponse (« prises » et « avec ») connaissaient réellement l'expression. L'item 13 a également connu un taux de succès de très faible (12,73 %), qui peut être attribuable à la faible fréquence du mot « lurette » dans les manuels scolaires (Manulex ne contient aucune occurrence de ce mot). De plus, la première syllabe de ce mot (« lu_____ ») a 138 concurrents dans Manulex : il est donc impossible que les enfants aient complété correctement l'expression sans la connaître. Enfin, ajoutons que cet item

comprenait deux mots à compléter, ce qui augmentait le niveau de difficulté. L'item 17 (flambant neuf) a lui aussi eu un taux de succès très faible 20,73 %). Signalons que le mot clé (« flambant ») est très rare (1 occurrence par million) et comporte un nombre important de concurrents dans Manulex (55). Outre la faible fréquence et les nombreux concurrents, l'insuccès des élèves à cet item peut être lié au fait que seules les lettres « fl » du mot ciblé (flambant) étaient fournies, contrairement aux autres items, où la première syllabe au complet était fournie (pour des raisons de phonétique telles qu'expliquées dans la *Méthodologie*). Par ailleurs, le mot ciblé était le premier mot de l'expression : si le mot tronqué avait été placé en deuxième position (ex. : flambant n____), l'expression aurait sans doute été plus facile à trouver : le mot « flambant » est fortement associé au mot « neuf », alors que le mot « neuf » peut être précédé d'une plus grande variété de mots (par exemple, les séquences « un vélo fleuri neuf » et « un vélo fluo neuf » comptent parmi les réponses des élèves ; ces syntagmes ne correspondent pas à la réponse attendue, mais ils sont sémantiquement plausibles et grammaticalement corrects). Quant à l'item 16 (sain et sauf), seulement 34,55 % des élèves l'ont complété correctement. Le faible taux de succès peut être attribuable au fait que le mot « sauf » est peu fréquent (2 occurrences par million) et qu'il a un nombre très élevé de concurrents (75). Cela dit, il est possible que certains enfants n'aient pas « reconnu » l'expression à partir du mot « sain », d'autant plus que ce mot comporte de nombreux homonymes (saint, ceint, sein, etc.), ce qui peut prêter à confusion. L'item 18 (à perte de vue) a été réussi par 35,76 % des élèves. Les deux mots ciblés avaient un nombre de concurrents assez élevé dans Manulex (15 pour « perte » et 66 pour « vue »). Il est donc improbable que des élèves puissent avoir deviné les mots de l'expression. L'item 15 (les quatre coins du monde) a également été réussi par une minorité d'élèves (35,76 %). Pourtant, les deux mots ciblés étaient très fréquents dans la base Manulex (443 et 565 occurrences respectivement) et avaient peu de concurrents (1 et 10 respectivement). On peut attribuer le faible taux de réussite au fait qu'il y avait 2 mots tronqués à compléter. De plus, en ce qui a trait au second mot cible « monde », seule la première lettre était fournie, ce qui augmentait la difficulté de reconnaître le mot.

5.5.2 L'effet du facteur « année scolaire » sur la réussite des items

Les analyses générales ont indiqué que l'effet « année scolaire » avait eu un effet significatif sur la performance des élèves de façon globale. Toutefois, ces analyses ne permettent pas de mesurer l'influence relative de ces facteurs sur la réussite de chaque item. C'est pourquoi nous proposons, dans cette sous-section, une analyse détaillée des items qui fera la lumière sur des questions plus spécifiques concernant l'effet de l'année scolaire : dans quels items peut-on observer une influence significative de l'année scolaire sur la réussite des élèves ? Ces items permettent-ils de distinguer tous les niveaux scolaires, ou certains niveaux seulement ?

En examinant le Tableau 5.5 (présenté plus haut), on peut constater que, pour la plupart des items, le taux de réussite a tendance à augmenter avec le niveau scolaire. Un test ANOVA a permis de mesurer l'effet du facteur « année scolaire » sur chaque item. Le test a révélé que le facteur année scolaire avait eu un effet significatif ($p < 0,05$) dans 14 items sur 18, soit les items 2 ($p=0,049$), 3 ($p<0,001$), 4 ($p<0,001$), 5 ($p=0,009$), 6 ($p<0,001$), 7 ($p<0,001$), 9 ($p<0,001$), 10 ($p<0,001$), 12 ($p=0,013$), 13 ($p=0,001$), 14 ($p<0,001$), 15 ($p=0,042$), 17 ($p=0,01$) et 18 ($p<0,001$). Ce facteur semble également avoir joué un rôle dans la réussite des items 1 et 16, pour lesquels la corrélation approchait le seuil significatif de $p = 0,05$ ($p = 0,068$ et $p = 0,069$, respectivement). Cela signifie que presque tous les items du test se sont avérés discriminants. Or, nous ne savons pas si ces items permettent de distinguer certains groupes seulement (par exemple les 3^e et le 6^e), ou tous les groupes (3^e, 4^e, 5^e, et 6^e). C'est pourquoi nous allons effectuer une analyse portant sur la progression par année, afin de déterminer précisément où se situent les écarts significatifs. Quant aux items pour lesquels l'année scolaire n'a pas eu d'effet significatif, ils feront également l'objet d'une brève discussion dans laquelle nous tenterons d'expliquer pourquoi ces items n'ont pas été discriminants.

A) Items reflétant une progression par année scolaire

Nous avons effectué une analyse post hoc de Tukey-Kramer : ce test a permis de faire des comparaisons deux à deux afin de savoir plus précisément où se situaient les différences entre les années scolaires. Il s'avère que certains items ont donné lieu à des écarts de performance significatifs entre plusieurs niveaux scolaires : c'est le cas des items 3, 4, 6, 7, 9, 10, 13, 14 et 18. D'autres items ont permis d'établir une distinction entre deux groupes seulement : il s'agit des items 5, 15 et 17. Cependant, les items 1, 2, 8, 11, 12, et 16 n'ont pas donné lieu à des différences significatives entre les niveaux scolaires.

À l'item 3 (plus ou moins), les élèves de 3^e année ont significativement moins bien performé que les élèves de 4^e année ($p = 0,049$), et que ceux de 5^e et de 6^e année ($p < 0,001$). Tel qu'expliqué précédemment, le fait que seules les premières lettres des mots cibles aient été fournies dans le test peut avoir posé des difficultés aux enfants, surtout les plus jeunes. Toutefois, il n'a pas permis de distinguer les élèves de 5^e et ceux de 6^e année : cela peut s'expliquer par le fait que cet item ne comportait que quelques concurrents plus fréquents (17); de plus, les deux mots à trouver («plus» et «moins») étaient des mots très fréquents. Quant à l'item 4 (en ce qui concerne), il a permis de distinguer les élèves de 3^e et de 6^e année : $3 < 6$ ($p = 0.007$), ainsi que les élèves de 4^e et 6^e année : $4 < 6$ ($p < 0,001$). Cet item a effectivement été réussi par seulement une minorité d'élèves de 3^e et de 4^e année ; il a toutefois été plutôt bien réussi par les 5^e et les 6^e années. Si cet item s'est révélé discriminant, c'est probablement en raison de la grande rareté du mot «concerne» (7,71 occurrences par million) et du nombre très élevé de concurrents plus fréquents (80). À l'item 5 (être à la hauteur), les élèves de 6^e année ont significativement mieux performé que les élèves de 3^e année : $3 < 6$ ($p = 0,03$). Pourtant, le mot clé «hauteur» est un mot familier (55 occurrences par million) et ne compte que deux (2) concurrents plus fréquents dans Manulex. Ensuite, l'item 6 (bel et bien) a donné lieu à des écarts significatifs entre les 3^e année et les autres groupes $3 < 4$ ($p = 0,019$), 5 ($p = 0,001$) et 6 ($p < 0,001$), de même qu'entre les 4^e et les 6^e année : $4 < 6$ ($p = 0,017$). Quant à l'item 7 (tout au long), les différences significatives se situent entre

les élèves de 3^e année et ceux de 5^e et de 6^e année $3 < 5$ ($p = 0,029$) et 6 ($p < 0,001$), ainsi qu'entre les élèves de 4^e et de 6^e année $4 < 6$ ($p = 0,024$). L'item 9 (de manière à ce que), quant à lui, a donné lieu à des différences significatives entre les 3^e et les 5^e année : $3 < 5$ ($p = 0,034$) ; entre les 3^e et les 6^e année : $3 < 6$ ($p < 0,001$), ainsi qu'entre les 4^e et les 6^e année : $4 < 6$ ($p = 0,006$). Ensuite, l'item 10 (au fur et à mesure) a mené à des écarts significatifs entre les élèves de 3^e et ceux de 5^e et de 6^e : $3 < 5$ ($p < 0,001$) et 6 ($p < 0,001$). À l'item 13 (depuis belle lurette), les élèves de 6^e année ont significativement mieux performé que ceux de 3^e année $3 < 6$ ($p = 0,015$) et que ceux de 4^e année $4 < 6$ ($p = 0,001$). En ce qui a trait à l'item 14 (à portée de la main), les élèves de 5^e et de 6^e année ont surpassé les élèves de 3^e année de façon significative : $3 < 5$ et 6 ($p < 0,001$). À l'item 15 (les quatre coins du monde), les élèves de 5^e année se sont démarqués par rapport à ceux de 3^e année (les quatre coins du monde) : $3 < 5$ ($p = 0,045$). À l'item 17 (flambant neuf), la différence significative se situe entre les 3^e et les 6^e année : $3 < 6$ ($p = 0,006$). Enfin, dans l'item 18 (à perte de vue), les élèves de 5^e et de 6^e année se sont distingués des élèves de 3^e année $3 < 5$ et 6 ($p < 0,001$) et des élèves de 4^e année : $4 < 5$ ($p = 0,037$) et 6 ($p < 0,001$).

À la lumière de ces données, nous pouvons remarquer que les écarts significatifs se situent généralement entre les élèves de 2^e cycle (3^e et 4^e) et ceux du 3^e cycle (5^e et 6^e). Aucun item n'a créé d'écart significatif entre les 5^e et les 6^e année, et seulement deux items (items 3 et 6) ont permis de distinguer les 3^e et les 4^e année.

B) Items ne reflétant pas de progrès significatif par année scolaire

Six items n'ont pas reflété de progression significative entre les années scolaires, soit les items 1 (à l'occasion de), 2 (tout le monde), 8 (être aux prises avec qqch), 11 (à tour de rôle), 12 (à toute vitesse) et 16 (sain et sauf). Certains items ont vraisemblablement été trop faciles pour la plupart des élèves, ce qui s'est traduit par une absence de progression significative d'une année à l'autre. C'est le cas des items 1 (à l'occasion de), 2 (tout le monde), 11 (à tour de rôle) et 12 (à toute vitesse).

Si l'on observe les taux de réussite de ces quatre items dans chaque année scolaire, on constate que les expressions ciblées sont connues de la majorité des élèves dès la 3^e année et que cette connaissance semble se maintenir (et dans certains cas, progresser légèrement) au fil des années. En comparant les taux de réussite des 3^e et des 6^e année aux items 2 et 12, nous remarquons une augmentation d'environ dix pour cent en faveur des élèves plus vieux ; en ce qui concerne l'item 1, cette augmentation dépasse les trente pour cent et frôle le seuil significatif ($p = 0,51$). Même si ces progressions ne sont pas statistiquement significatives, nous pouvons néanmoins affirmer que la proportion d'élèves ayant réussi ces items a tendance à augmenter en fonction des niveaux scolaires, du moins entre la 3^e et la 6^e année, et particulièrement à l'item 1. Il est donc possible que l'année scolaire ait exercé une influence sur la réussite de ces items.

Cela dit, d'autres items ont au contraire été trop difficiles pour la majorité des élèves, ce qui s'est également reflété par une absence de progression significative. C'est le cas des items 8 (être aux prises avec qqch) et 16 (sain et sauf). L'item 8 est, rappelons-le, le moins bien réussi par l'ensemble des élèves (taux de réussite de 4%); il s'agit également de l'item affichant la plus faible progression de taux de réussite (celui passe de 0 % en 3^e année à 6,5 % en 6^e année). Visiblement, cette expression n'est pas à la portée des élèves du primaire. Quant à l'item 16 (sain et sauf), il semble aussi avoir occasionné des difficultés aux élèves de tous les niveaux. Si le taux de réussite tend à augmenter avec le niveau scolaire (voir Tableau 5.5), il ne s'agit pourtant pas d'un progrès significatif.

5.5.3 Items reflétant une influence du facteur « groupe linguistique »

Tout d'abord, rappelons que les tests statistiques n'ont pas révélé d'effet significatif du groupe linguistique sur la réussite de l'ensemble des items. Or, il se peut que ce facteur ait créé des écarts de performance significatifs entre les groupes linguistiques dans certains items. Nous allons donc examiner l'influence possible du facteur « langue » sur la réussite des items dans chaque année scolaire.

Chez les élèves de 3^e année, le facteur « langue » n'a pas exercé d'influence significative sur la réussite des items. Toutefois, on peut observer une différence

presque significative à l'item 3 (plus ou moins) en faveur des allophones : ($p = 0,058$), ainsi qu'à l'item 9 en faveur des francophones (de manière à ce que) : ($p = 0,067$). Chez les élèves de 4^e année, on peut observer une influence significative du facteur « langue » dans l'item 5 (être à la hauteur) : les francophones ont mieux réussi que les allophones ($p = 0,035$). Chez les élèves de 5^e année, les allophones ont mieux performé que les francophones à l'item 3 (plus ou moins) : ($p = 0,018$), à l'item 12 (à toute vitesse) : ($p = 0,041$) et à l'item 13 (depuis belle lurette) : ($p = 0,041$). En ce qui concerne les 6^e année, le facteur « langue » n'a produit aucun effet significatif sur les items. Enfin, dans le groupe DGA, nous pouvons constater que les allophones ont mieux réussi que les francophones à l'item 7 ($p=0,003$) et à l'item 9 ($p=0,011$).

De prime abord, nous pouvons constater que parmi les items où le facteur « langue » a exercé une influence significative, ce sont les allophones qui ont le mieux réussi (dans 5 cas sur 6). D'ailleurs, si l'on observe le Tableau 5.2, on peut remarquer un léger avantage en faveur des allophones en ce qui a trait à la proportion d'items réussis, quoique cet écart ne soit pas significatif. Donc, bien qu'il n'y ait pas eu de différence significative entre les deux groupes linguistiques de façon globale, nous pouvons constater une avance de la part des allophones dans certains items. Cela dit, il est intéressant de noter que la plupart des items qui ont engendré une différence significative entre les allophones et les francophones sont des items qui ont reflété une progression par année scolaire (soit les items 3, 5, 7, 9, et 13). Toutefois, force est de reconnaître que les écarts observés entre les deux groupes linguistiques constituent des exceptions : la grande majorité des items ne reflètent aucune influence du facteur « langue ».

5.6 Conclusion

En début de chapitre, nous avons vérifié si l'année scolaire avait eu un effet sur la réussite des élèves. Notre première hypothèse était que les élèves plus avancés dans leur scolarisation avaient des connaissances plus solides et plus étendues en matière d'expressions figées que les élèves moins avancés. Nous nous attendions

donc à ce que les élèves plus avancés performant mieux au test que les élèves moins avancés. Les analyses statistiques que nous avons réalisées ont permis de confirmer notre hypothèse. En effet, les élèves de 3^e année ont moins bien performé que ceux de 4^e ($p=0,0272$), et que ceux de 5^e et de 6^e année ($p<0,0001$). De plus, les élèves 4^e année ont moins bien réussi que les élèves de 5^e ($p=0,0287$) et de 6^e ($p<0,0001$). Ensuite, nous avons mesuré l'influence du groupe linguistique sur la performance des participants. Notre hypothèse était que les élèves francophones obtiendraient une meilleure performance que les élèves allophones en raison de leur exposition accrue au français parlé à l'extérieur du milieu scolaire. Cependant, nos analyses n'ont révélé aucune différence significative entre ces deux groupes linguistiques. Nous avons aussi exploré la possibilité d'un écart de performance entre les filles et les garçons ; toutefois, le facteur sexe n'a pas joué de rôle significatif dans la réussite des élèves. Par la suite, nous avons exploré l'influence de trois facteurs linguistiques sur la réussite des élèves, soit la fréquence des expressions ciblées, la fréquence des mots clés et le nombre de concurrents plus fréquents des mots clés. La fréquence des expressions ne semble pas avoir joué de rôle significatif dans le succès des élèves. Par contre, la fréquence des mots clés, ainsi que le nombre de concurrents ont exercé une influence significative dans presque tous les groupes et sous-groupes étudiés. Des explications ont été proposées, mais nous considérons que ce sujet mériterait une attention plus particulière lors de recherches futures. Nous avons ensuite poursuivi notre discussion avec une analyse par item. Nous avons d'abord étudié la performance générale des élèves à chaque item, et nous avons tenté d'expliquer les taux de succès en fonction de la fréquence des mots clés et du nombre de concurrents plus fréquents. Nous nous sommes ensuite penchée sur les items qui reflétaient une progression par année scolaire. En effet, la plupart des items se sont avérés discriminants, et plus particulièrement entre les élèves de 2^e et de 3^e cycle. La discussion a ensuite été axée sur l'effet possible du groupe linguistique sur le succès des élèves aux différents items. Nous avons découvert que l'effet « langue » semblait avoir eu un effet significatif dans quelques items. Les résultats ayant été présentés et interprétés, nous allons maintenant passer à la conclusion de notre

recherche. Celle-ci constituera un bref bilan de notre démarche et de nos résultats ; nous y proposerons également quelques pistes de réflexion sur le rôle des expressions figées dans l'apprentissage et la connaissance des langues.

CONCLUSION

L'objectif de cette étude était de mesurer l'influence de plusieurs facteurs sur la connaissance des expressions figées chez des enfants de niveau primaire. Les facteurs étudiés sont les suivants : le niveau scolaire, le groupe linguistique, le sexe et les caractéristiques linguistiques (la fréquence des expressions, la fréquence des mots clés et le nombre de concurrents plus fréquents que les mots clés). Afin de répondre à nos questions de recherche, nous avons créé un test de production guidée que nous avons administré à des élèves d'une école primaire de Montréal. Notre population était formée d'élèves allophones et francophones répartis sur quatre niveaux scolaires : 3^e, 4^e, 5^e et 6^e année. Les groupes linguistiques ont été établis en fonction de la langue parlée le plus souvent à la maison (avec les parents, plus précisément). À partir des données recueillies, nous avons effectué une série d'analyses statistiques afin de mesurer l'impact de chacun des facteurs mentionnés sur la performance des élèves. À la lumière de ces analyses, nous avons d'abord constaté que le niveau scolaire avait fortement influencé la performance des élèves. En effet, les élèves des niveaux scolaires plus avancés ont mieux performé que les élèves des niveaux moins avancés, et cette tendance s'est révélée statistiquement significative. Ces résultats portent à croire que plus les enfants avancent dans leur scolarisation, plus ils développent les capacités nécessaires à la production d'expressions figées.

En ce qui concerne l'influence du groupe linguistique, ce facteur n'a pas donné lieu à une différence significative entre les performances des allophones et des francophones de façon globale. Nous avons néanmoins constaté quelques différences de performance au niveau des items, mais ces écarts constituent des

exceptions. Ce résultat ne concorde pas avec l'hypothèse que nous avons formulée : nous nous attendions à ce que les francophones réussissent mieux que les allophones, ces derniers étant moins exposés au français parlé que les autres élèves dans le milieu familial. Ce résultat inattendu peut être attribuable au fait que presque tous les élèves évalués, y compris les allophones, fréquentaient une classe régulière francophone depuis la maternelle. Les élèves allophones, en plus d'avoir bénéficié d'une quantité appréciable d'input oral et écrit dès le préscolaire, ont appris le français dans une situation authentique de communication, ce qui semble avoir favorisé la capacité de comprendre et d'employer des expressions figées. Dans un autre ordre d'idées, les renseignements recueillis auprès des élèves ne nous ont pas permis de tenir compte des situations linguistiques parfois complexes dans les familles des élèves allophones. En effet, nous avons pris en considération la langue parlée le plus souvent avec les parents; or, vu le nombre d'élèves et les contraintes de temps, il nous était impossible de contrôler avec précision l'exposition réelle des jeunes allophones au français parlé. Nous estimons que cet aspect mériterait davantage d'attention lors de recherches impliquant un nombre plus restreint de participants.

L'effet possible du facteur « sexe » a également été exploré. Nous avons effectivement envisagé la possibilité que les filles performant mieux que les garçons, ces dernières obtenant généralement de meilleurs résultats en français que les garçons. Or, nos analyses n'ont pas révélé d'écart significatif entre ces deux groupes. Il se peut que le facteur « sexe » ait peu ou pas d'influence sur la connaissance et l'emploi d'expressions figées pour l'ensemble des locuteurs. Il est aussi possible que ce facteur exerce une influence pendant des périodes de la vie autres que celle que nous avons étudiée (préscolaire, adolescence, âge adulte).

Quant aux facteurs linguistiques étudiés, deux d'entre eux ont joué un rôle significatif sur la réussite des élèves, soit la fréquence des mots clés et le nombre de concurrents plus fréquents. En effet, les enfants ont eu davantage de succès aux items dont les mots clés (mots à compléter) étaient fréquents qu'aux items dont les mots clés étaient rares. Ils ont aussi mieux réussi les items dont les mots clés avaient peu de concurrents que les items dont les mots clés avaient un nombre

élevé de concurrents. Nos analyses statistiques ont confirmé que ces deux facteurs avaient eu un effet significatif sur la performance des participants. Par contre, la fréquence des expressions ne semble pas avoir influencé la réussite des élèves, contrairement à nos attentes. Il est possible que la base de données utilisée pour mesurer la fréquence des expressions soit en cause : elle reflète le lexique auquel les adultes, mais pas nécessairement les enfants, sont exposés.

Cela dit, bien que toute notre démarche ait été effectuée dans un souci de rigueur scientifique, nous reconnaissons que la présente étude comporte des limites. Puisqu'il s'agit d'une étude exploratoire et de portée assez restreinte, les résultats obtenus doivent être interprétés avec prudence. Soulignons d'abord que les élèves ont été évalués à l'aide d'un seul test consistant en une tâche de production guidée. Ce test n'a permis de mesurer qu'une partie des compétences réelles des élèves en matière d'expressions figées : les résultats auraient pu se révéler différents si l'on avait eu recours à un ou plusieurs autres instruments de mesure complémentaires, comme l'on fait Schmitt et ses collègues (2004). Par ailleurs, la capacité d'un élève de produire correctement des expressions figées en situation de production guidée n'implique pas nécessairement qu'il soit capable de les employer spontanément et correctement dans un contexte de production libre. Néanmoins, nous croyons que les résultats obtenus à l'aide de notre instrument de mesure constituent un bon indicateur de la capacité des élèves à employer des EF dans leur production langagière : « although the ability to complete a cloze item successfully does not demonstrate the ability to use the formulaic sequence at will in discourse, it does give persuasive evidence of at least some degree of productive knowledge » (Schmitt *et al.* 2004, p. 67). Cela dit, rappelons que le test comportait seulement 18 items, ce qui ne représente qu'un faible échantillon d'expressions figées. Enfin, il faut reconnaître qu'une collecte de données effectuée en milieu scolaire comporte nécessairement des contraintes (évaluation en groupe, distractions, limites de temps, etc.). Néanmoins, nous sommes d'avis que malgré quelques inconvénients, une collecte de données effectuée en classe présente l'avantage de créer des conditions qui se rapprochent de situations authentiques d'évaluation. Ainsi, nous croyons que

la performance des élèves doit refléter de façon assez réaliste leurs connaissances des expressions qui figuraient dans le test.

Cette recherche n'a pas pu aborder certaines questions. Plutôt que d'y voir une lacune, nous y voyons plutôt l'occasion d'approfondir la question dans des recherches ultérieures. En outre, il serait intéressant de mesurer la performance d'une population d'élèves à l'aide d'un test comportant un plus grand nombre et une plus grande variété d'expressions figées, en incluant par exemple des collocations et des expressions idiomatiques. Une telle recherche pourrait servir, d'une part, à identifier les difficultés propres aux divers types d'expressions, et d'autre part, à mettre au point une estimation du nombre d'expressions figées connues à chaque année scolaire. Il serait également souhaitable d'évaluer les élèves à l'aide d'une batterie complète de tests sur les formes figées : tests à choix multiples, repérage d'expressions figées dans des textes, entrevues, productions écrites. Un tel ensemble de tests permettrait de mesurer à la fois les connaissances réceptives et productives, tant à l'oral qu'à l'écrit. De plus, il serait pertinent de vérifier l'influence possible d'autres caractéristiques linguistiques sur la connaissance des expressions figées. Beaucoup de chercheurs se sont intéressés à l'effet de la fréquence et du degré d'opacité sémantique, mais d'autres facteurs, comme le degré de figement, la syntaxe marquée et la présence de mots à contexte unique, sont aussi susceptibles d'avoir un effet sur l'acquisition des EF, surtout chez les enfants et les chez les locuteurs non natifs.

Dans un autre ordre d'idées, la présente étude suscite également des questions importantes concernant l'évolution du langage créatif (non figé) chez les élèves du primaire. En effet, notre recherche a permis de constater que chez les enfants allophones et francophones, la maîtrise des expressions figées semble se renforcer d'année en année. Or, nous ne savons pas dans quelle mesure les compétences des enfants en matière de combinaisons lexicales libres évoluent de la 3^e à la 6^e année. Tel que mentionné dans le *Cadre théorique*, la proportion de langage figé et de langage créatif varie pendant l'enfance et l'adolescence ; ce n'est que vers l'âge de 18 ans que s'établit l'équilibre entre l'utilisation du langage créatif et du langage figé (Wray 2002, p. 40). Si l'on mettait au point une batterie de tests permettant de

mesurer, dans un groupe d'élèves donné, à la fois les habiletés relatives au langage figé ainsi que les aptitudes ayant trait au langage créatif, il serait alors possible de voir de quelle façon ces deux axes de compétences se développent et interagissent au fil des ans.

Cette recherche soulève également des questions importantes concernant l'acquisition du français chez les jeunes allophones montréalais. Nos résultats semblent indiquer que les élèves allophones scolarisés en français depuis la maternelle aient des compétences comparables à celles de leurs camarades francophones dans le cadre d'une tâche de production écrite guidée. Bien que ce constat puisse paraître encourageant pour les jeunes allophones de cette école, nous croyons nécessaire de poursuivre les recherches sur l'acquisition des expressions figées chez les allophones : la présente étude, de nature exploratoire, ne donne qu'un aperçu des compétences d'un groupe d'allophones en matière de langage figé. Afin de bien cerner les connaissances des allophones en la matière, il faudrait en outre se pencher sur le cas des allophones qui intègrent le réseau scolaire francophone plus tardivement (au cours de l'école primaire ou secondaire, par exemple). Ces derniers sont plus susceptibles de vivre des difficultés relatives à l'acquisition et à l'emploi d'expressions figées; en mettant au point une batterie de tests appropriée, on serait davantage en mesure de détecter les faiblesses possibles et d'y remédier. Il serait aussi important d'analyser en profondeur les productions écrites et orales des allophones; nous croyons que les productions libres refléteraient davantage la capacité réelle des allophones à employer du langage figé. Une telle démarche permettrait notamment de mesurer la proportion de langage figé et de langage non figé, de vérifier quels types d'expressions sont les mieux maîtrisés (idiomatiques, collocations, locutions, etc.), et d'évaluer le niveau d'idiomaticité du discours oral et écrit des apprenants. Des études de corpus ont été réalisées entre autres par Forsberg (2006) et Nesselhauf (2005) chez des apprenants adultes LNN, mais à notre connaissance, il n'existe pas encore de recherche de cette envergure concernant le langage figé chez les enfants allophones. Enfin, il serait aussi intéressant de vérifier s'il existe un lien entre le niveau de compétence des enfants allophones en matière d'expressions figées et les autres compétences linguistiques,

comme la compréhension en lecture, l'étendue du lexique et la maîtrise des règles grammaticales.

En terminant, nous tenons à souligner de nouveau l'importance de l'utilisation du langage figé dans l'apprentissage d'une langue et dans la communication. En effet, il a été démontré à maintes reprises que la maîtrise d'un grand nombre d'expressions figées avait un effet positif sur la capacité de s'exprimer de façon fluide et idiomatique, tant chez les locuteurs natifs que non natifs. L'apprentissage et l'utilisation d'une grande variété d'expressions figées devraient donc être une préoccupation majeure en didactique des langues, non seulement chez les professeurs, mais aussi chez les apprenants aspirant à un haut niveau de compétence dans la langue cible.

BIBLIOGRAPHIE

- Abel, Beate. 2003. «English idioms in the first language and second language lexicon : a dual representation approach». *Second Language Research*, vol.19, n° 4, p. 329-358.
- Ackerman, Brian P. 1982. «On comprehending idioms: Do children get the picture?». *Journal of Experimental Psychology*, vol. 33, n° 3, p. 439-454.
- Bartning, Inge et Fanny Forsberg. 2006. «Les séquences préfabriquées à travers les stades de développement en français L2». *Actes du XVIe congrès des romanistes scandinaves*. <http://www.ruc.dk/cuid/publikationer/publikationer/XVI-SRK-Pub/TVI/TVI18-BartningForsberg/>.
- Bee, Helen et Denise Boyd. 2008. *Les âges de la vie: psychologie du développement humain*. Saint-Laurent : ERPI, 468 p.
- Blaye, Agnès et Patrick Lemaire. 2007. *Psychologie du développement cognitif de l'enfant*. Bruxelles: De Boeck, 430 p.
- Beer-Toker, Mia et Andrée Gaudreau. 2006. «Représentations, attitudes et pratiques de littératie chez les élèves allophones : construction d'un outil de dépistage des difficultés en matière de littératie». *Revue des sciences de l'éducation*, vol. 32, n°2, p. 345-376.
- Bernicot, Josée. 2000. «La pragmatique des énoncés chez l'enfant». Chap. in *L'acquisition du langage*, tome 2, sous la dir. de M. Kail et M. Fayol, p. 45-82. Paris : Presses Universitaires de France.
- Butler, Christopher S. 2004. «Corpus studies and functional linguistic theories». *Functions of Language*, vol. 11, n° 2, p. 147-186.

- Bybee, Joan. 2002. «Phonological evidence for exemplar storage of multiword sequences». *Studies in Second Language Acquisition* vol. 24, n° 2, p. 215-221.
- Cacciari, Cristina et Patrizia Tabossi. 1988. «The comprehension of idioms». *Journal of Memory and Language*, vol. 27, n° 6, p. 688-683.
- Chanier, Thierry. 1997. «Alexia : un environnement d'aide à l'apprentissage lexical du français langue seconde». Chap. in : *La locution : entre lexique, syntaxe et pragmatique. Identification en corpus, traitement, apprentissage*, sous la dir. de Pierre Fiala, Pierre Lafon, Marie-France Piguet, p. 105-118. Paris : INALF.
- Chenoweth, Ann N. 1995. «Formulaicity in essay exam answers». *Language Sciences*, vol.17, n° 3, p. 283-297.
- Clark, Ruth. 1974. «Performing without competence». *Journal of Child Language*, vol.1, n° 1, p. 1-10.
- Clark, Ruth. 1977. «What's the use of imitation?», *Journal of Child Language*, vol. 4, n°3, p. 341-358.
- Conklin, Kathy et Norbert Schmitt. 2007. «Formulaic Sequences: Are They Processed More Quickly than Nonformulaic Language by Native and Nonnative Speakers?» *Applied Linguistics*, vol. 29, n° 1, p. 72-89.
- Denhière, Guy et Jean-Claude Versteig. 1997. «Le traitement cognitif des expressions idiomatiques : activités automatiques et délibérées». Chap. in : *La locution : entre lexique, syntaxe et pragmatique. Identification en corpus, traitement, apprentissage*, sous la dir. de Pierre Fiala, Pierre Lafon, Marie-France Piguet, p. 119-148. Paris : INALF.
- Dörnyei, Zoltan, Valerie Durow et Khawla Zahran. 2004. «Individual differences and their effects on formulaic sequence acquisition». Chap. in *Formulaic sequences: Acquisition, processing and use*, sous la dir. de Norbert Schmitt, p. 87-106. Philadelphie : John Benjamins.
- Ellis, Nick C. 2002. «Frequency effects in language processing; A review with implications for theories of implicit and explicit language acquisition». In *Studies in Second Language Acquisition*, vol. 24, n° 2, p. 143-188.

- Ellis, Nick C. 2003. «Constructions, Chunking, and Connectionism: The Emergence of Second Language Structure». Chap. in: *The Handbook of Second Language Acquisition*, sous la dir. de C. Doughty et M. Long, p. 63-103. Victoria, Australie : Blackwell Publishing.
- Ellis, Rod. 1984. *Classroom Second Language Development*. Oxford : Pergamon Institute of English, 322 p.
- Erman, Britt et Beatrice Warren. 2000. «The idiom principle and the open choice principle». *Text*, vol. 20, n°1, p. 29-62.
- Fitzpatrick, Tess et Alison Wray. 2006. «Breaking up is not so hard to do: individual differences in L2 memorization». *Canadian Modern Language Review*, vol. 63, n°1, p. 35-57.
- Forsberg, Fanny. 2006. « Le langage préfabriqué en français parlé L2 ». Thèse de doctorat, Stockholm, Université de Stockholm, 183 p.
- Foster, Pauline. 2001. «Rules and routines : a consideration of their role in the task-based language production of native and non-native Speakers». Chap. in *Researching pedagogic tasks : second language learning, teaching and testing*, sous la dir. de M. Bygate, P. Skehan et M. Swain, p. 75-94. Londres et New York : Longman.
- Gibbs, Raymond W. 1987. «Linguistic factors in children's understanding of idioms». *Journal of Child Language*, vol.14, n° 3, p. 569-586.
- Gross, Gaston. 1996. *Les expressions figées en français. Noms composés et autres locutions*. Paris : Ophrys, 161p.
- Hoey, Michael. 2005. *Lexical Priming: A New Theory of Words and Language*. New York : Routledge, 201 p.
- Iralde, Lydie et Aurélie Lainé. 2004. «Compréhension d'expressions idiomatiques familières selon leur degré de transparence métaphorique chez l'enfant et l'adolescent de langue maternelle française». *Le langage et l'homme*, vol. 39, n° 2, p. 141-166.
- Jiang, Nan et Tatiana M. Nekrasova. 2007. «The Processing of Formulaic Sequences by Second Language Speakers». *The Modern Language Journal*, vol. 91, n° 3, p. 433-445.

- Krashen, Stephen D. et Robin Scarcella. 1978. «On routines and patterns in language acquisition and performance». *Language Learning*, vol. 28, n° 2, p. 283-300.
- Laval, Virginie. 2003. «Idiom comprehension and metapragmatic knowledge in French children». *Journal of Pragmatics*, vol. 35, n° 5, p.723-739.
- Lété, Bernard. 2004. «Manulex : une base de données du lexique écrit adressé aux élèves». Chap. in : *Didactique du lexique : contextes, démarches, supports*, sous la dir. d'E. Calaque et J. David. Bruxelles, p. 241-257. Louvain-la-Neuve, Belgique : Éditions de Boeck.
- Levorato, Maria Chiara et Cristina Cacciari. 1995. «The effects of different tasks on the comprehension and production of idioms in children». *Journal of experimental child psychology*, vol. 60, n° 2, p. 261-283.
- Levorato, Maria Chiara et Cristina Cacciari. 1992. «Children's comprehension and production of idioms: the role of context and familiarity». *Journal of Child Language*, vol. 19, n° 2, p. 415-433.
- Mel'čuk, Igor. 1993. «La phraséologie et son rôle dans l'enseignement/apprentissage d'une langue étrangère.» *Étude de Linguistique Appliquée*, vol. 92, n° 4, p. 82-113.
- Myles, Florence, Janet Hooper et Rosamond Mitchell.1998. «Rote or Rule? Exploring the Role of Formulaic Language in Classroom Foreign Language Learning». *Language Learning*, vol. 48, n°3, p. 323-363.
- Myles, Florence, Janet Hooper et Rosamond Mitchell.1999. «Interrogative Chunks in French L2». *Studies in Second Language Acquisition*, vol. 21, n° 1, p. 49-80.
- Nattinger, James R. et Jeannette S. DeCarrico. 1992. *Lexical Phrases and Language Teaching*, Oxford : Oxford University Press, 218 p.
- Nesselhauf, Nadja. 2005. *Collocations in a Learner Corpus*. Studies in Corpus Linguistics. Amsterdam : John Benjamins, 329 p.
- Nippold, Marilyn A. et Mishelle Rudzinski. 1993. «Familiarity and transparency in idiom explanation: A developmental study of children and adolescents». *Journal of Speech and Hearing Research*, vol. 36, n° 4, p. 728-737.

- Nippold, Marilyn A. et Catherine L. Taylor. 1995. «Idiom understanding in youth: further examination of familiarity and transparency». *Journal of Speech & Hearing Research*, vol. 38, n° 2, p. 426-433.
- Nippold, Marilyn A. et Catherine L. Taylor. 2002. «Judgments of idiom familiarity and transparency: a comparison of children and adolescents». *Journal of Speech & Hearing Research*, vol. 45, n° 2, p. 384-391.
- Oppenheim, Nancy. 2000. «The Importance of Recurrent Sequences for Nonnative Speaker Fluency and Cognition». Chap. in : *Perspectives on Fluency*, sous la dir. de Heidi Riggenbach, p. 220-240. Ann Arbor : University of Michigan Press.
- Pawley, Andrew et Frances H. Syder. 2000. «The One-Clause-at-a-Time-Hypothesis». Chap. in : *Perspectives on Fluency*, sous la dir. de Heidi Riggenbach, p. 163-199. Ann Arbor : University of Michigan Press.
- Pawley Andrew et Frances H. Syder. 1983. «Two puzzles for linguistic theory: nativelike selection and nativelike fluency». Chap. in : *Language and Communication*, p. 191-226. New York : Longman.
- Perera, Natsuko. 2001. «The Role of Prefabricated Language in Young Children's Second Language Acquisition». *Bilingual Research Journal*. vol. 25, n° 3, p. 327-356.
- Perkins, Michael R. 1999. «Productivity and formulaicity in language development». Chap. in *Issues in normal & disordered child language: from phonology to narrative*, sous la dir. de M. Garman, C. Letts, B. Richards, C. Schellletter & S. Edwards, p. 51-67. Édition spéciale de Bulmershe Papers. Reading: University of Reading.
- Peters, Ann M. 1983. *The units of language acquisition*. Cambridge, Angleterre : Cambridge University Press, 131 p.
- Pulido, Loïc, Lydie Iralde et Annick Weil-Barais. 2007. «La compréhension des expressions idiomatiques à 5 ans : une étude exploratoire». *Enfance*, vol. 59, n° 4, p. 339-355.
- Raupach, Manfred. 1984. «Formulae in second language speech production». Chap. in *Second Language Production*, sous la dir. de Hans W. Dechert, Dorothea Möhle et Manfred Raupach, p. 114-137. Tübingen, Allemagne : Gunter Narr.

- Towell, Richard, Roger Hawkins et Nives Bazergui. 1996. «The development of fluency in advanced learners of French». *Applied Linguistics*, vol. 17, no 1, p. 84-119.
- Schapira, Charlotte. 1999. *Les stéréotypes en français : proverbes et autres formules*. Paris : Ophrys, 172 p.
- Schmitt, Norbert et Ronald Carter. 2004. «Formulaic sequences in action: an introduction». Chap. in *Formulaic Sequences: Acquisition, processing and use*, sous la dir. de Norbert Schmitt, p. 1-22. Philadelphie : John Benjamins.
- Schmitt, Norbert, Zoltan Dörnyei, Svenja Adolphs et Valerie Durow. 2004. «Knowledge and acquisition of formulaic sequences». Chap. in *Formulaic Sequences : Acquisition, processing and use*, sous la dir. de Norbert Schmitt, p. 55-86. Philadelphie : John Benjamins.
- Sinclair, John. 1991. *Corpus, Concordance and Collocation*. Oxford: Oxford University Press, 197 p.
- Spöttl, Carol et Michael McCarthy. 2003. «Formulaic utterances in the multilingual context». Chap. in *The Multilingual Lexicon*, sous la dir. de Jasone Cenoz, Britta Hufeisen et Ulrike Jessner, p. 133-151. Dordrecht : Kluwer Academic Publishers.
- Svensson, Maria Helena. 2004. «Critères de figement : l'identification des expressions figées en français contemporain». Thèse de doctorat, Umeå, Université d'Umeå, 193 p.
- Titone, Debra A. et Cynthia M. Connine. 1994. «Comprehension of idiomatic expressions : effects of predictability and literality». *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 20, n° 5, p. 1126-1138.
- Underwood, Geoffrey, Norbert Schmitt et Adam Galpin. 2004. «The eyes have it: An eye-movement study into the processing of formulaic sequences». Chap. in *Formulaic Sequences : Acquisition, processing and use*, sous la dir. de Norbert Schmitt, p. 153-168. Philadelphie : John Benjamins.
- Weinert, Regina. 1995. «The Role of Formulaic Language in Second Language Acquisition: A Review». In: *Applied Linguistics*, vol.16, n° 2, p. 180-205.

- Wong-Fillmore, Lily M. 1976. «The second time around: cognitive and social strategies in second language acquisition». Thèse de doctorat. Stanford, Université de Stanford.
- Wood, David. 2001. «In Search of Fluency: What Is It and How Can We Teach It?». *Canadian Modern Language Review*, vol. 57, n° 4, p. 573-589.
- Wood, David. 2002. «Formulaic Language in Acquisition and Production: Implications for Teaching». *Revue TESL du Canada*, vol. 20, n°1, p. 1-15.
- Wood, David. 2006. «Uses and Functions of Formulaic Sequences in Second Language Speech : An Exploration of the Foundations of Fluency». *The Canadian Modern Language Review*, vol. 63, n° 1, p.13-33.
- Wray, Alison. 2007. *Formulaic Language : Pushing the Boudaries*. Manuscrit.
- Wray, Alison. 2002. *Formulaic Language and the Lexicon*. Cambridge, Angleterre : Cambridge University Press, 332 p.
- Wray, Alison. 2000. «Formulaic Sequences in Second Language Teaching: Principle and Practice». *Applied Linguistics*, vol. 21, n° 4, p. 463-489.
- Wray, Alison. 1999. «Formulaic language in learners and native Speakers». *Language Teaching*, vol. 32, n° 4, p. 213-231.
- Wray, Alison et Michael R. Perkins. 2000. «The functions of formulaic language: an integrated model». *Language & Communication*, vol. 20, n° 1, p. 1-28.

Ouvrages de référence

- Beauchesne, Jacques. 2001. *Dictionnaire des cooccurrences*. Montréal : Éditions Guérin.
- De Villers, Marie-Éva. 2003. *Multidictionnaire de la langue française*. Montréal : Québec Amérique.
- Rey, Alain et Chantreau, Sophie. 2003. *Dictionnaire d'expressions et locutions*. Paris: Le Robert.

Publications gouvernementales

Ottawa, Ontario. Statistique Canada. Ressources humaines et développement social Canada. 2008. *Les compétences en lecture des jeunes immigrants au Canada : les effets de la durée de résidence, de l'exposition aux langues parlées à domicile, et des écoles*. Juin 2008. Accessible sur internet :

http://www.rhdsc.qc.ca/fr/publications_ressources/politique_sur_apprentissage/sp_849_06_08/page05.shtml

Ottawa, Ontario. Statistique Canada. Études sur la famille et le travail. 2001. *Le rendement scolaire des enfants d'immigrants au Canada, 1994-98*. Christopher Worswick. Rapport n° 178.

Ottawa, Ontario. Statistique Canada. 2006. *Titre du tableau : langue maternelle détaillée (103), langue parlée le plus souvent à la maison (8), autre langue parlée régulièrement à la maison (9) et groupes d'âge (17a) pour la population, pour le Canada, les provinces, les territoires, les régions métropolitaines de recensement et les agglomérations de recensement. Recensement de 2006*. Produit n° 97-555-XCB2006028 au catalogue de Statistique Canada.

http://cansim2.statcan.ca/cgiwin/cnsmcgi.pgm?Lang=F&EF_Action=Result&EF_ID=50006&EF_TYP=60&EF_Sort=-2

Québec, ministère de l'Éducation. *La réussite des garçons, des constats à mettre en perspective*. Rapport synthèse. Accessible sur internet : http://www.mels.gouv.qc.ca/stat/recherche/reussite_garcon.pdf, consulté le 15/08/2008.

Québec, ministère de l'Éducation, du Loisir et du Sport. Site «Livres ouverts», accessible à l'adresse suivante : http://www.livresouverts.qc.ca/index_V2.php, consulté le 26/03/2007.

Québec, ministère de l'Immigration et des Communautés culturelles. Statistiques du 3^e trimestre de 2007. Site internet : <http://www.micc.gouv.qc.ca/publications/fr/recherches-statistiques/BulletinStatistique-2007trimestre3-ImmigrationQuebec.pdf>. Consulté le 05/02/2008.

Sites internet consultés

Commission scolaire de Montréal
<http://www.csdm.qc.ca/Csdm/Administration/default.aEF?csdm=mosaique>

Fédération des commissions scolaires du Québec

<http://www.fcsq.qc.ca/>

Lexique

<http://www.lexique.org>

Office québécois de la langue française. Grand dictionnaire terminologique.

<http://www.granddictionnaire.com>

Manuels scolaires consultés

Deuxième cycle du primaire (3^e et 4^e année)

Bédard, Raymond, Guay, Luc et Diotte, Carl. 2002. *Paysages d'ici et d'ailleurs: géographie, histoire, éducation à la citoyenneté*, Manuel A et Manuel B. Laval : Éditions HRW.

Belley, Guylène. 2000-2001. *Aventure : expéditions au cœur de la science et de la technologie : primaire, 2^e cycle*. Tome 1 et Tome 2. Montréal : Lidec.

Bernier, Brigitte. 2002. *Sur la piste, 2^e cycle du primaire : géographie, histoire et éducation à la citoyenneté*. Tome A et Tome B. Saint-Laurent : Éditions du Renouveau pédagogique.

Bettinger, Simone. 2002. *Tangram*. Manuel C. Saint-Laurent : Éditions du Renouveau pédagogique.

Roy, Alain. 2001. *L'odyssée de la science et de la technologie, 2^e cycle*. Tome A et Tome B. Anjou : Éditions CEC.

Trudeau, Sophie et Létourneau, Ginette. 2002. *Mes chantiers*. Manuel A et Manuel B. Boucherville, Québec : Graficor.

Troisième cycle (5^e et 6^e année)

Bernier, Brigitte. 2003. *Sur la piste, 3^e cycle du primaire : géographie, histoire et éducation à la citoyenneté*. Manuel A et Manuel B. Saint-Laurent, Québec : ERPI.

Dulude, Françoise. 2003-2004. *Signet : français, 3^e cycle du primaire*. Livre A. Saint-Laurent, Québec : Éditions du Renouveau pédagogique.

Jean, Gaëtan. 2004. *Voyages : géographie, histoire et éducation à la citoyenneté : 3^e cycle du primaire*. Manuel A et Manuel B. Anjou, Québec : Éditions CEC.

Manuels scolaires, premier cycle du secondaire (Secondaire 1, 2 et 3)

Bergeron-Tremblay, Mariette. 1996. *Mots de passe : première secondaire*. Laval : Mondia.

Paradis, Odile, Marcel Poirier et Laurent Saint-Pierre. 1996. *Écologie : un monde à découvrir : 1re secondaire*. 2e édition. Montréal : HRW.

Rousselle, James. 1998. *Pour lire et pour écrire : français, deuxième secondaire*. Montréal : Centre éducatif et culturel.

Littérature jeunesse (romans)

Anfousse, Ginette. 1990. *Les vacances de Rosalie*. Montréal : La courte échelle (collection Jeunesse), 92 p.

Décary, Marie. 1993. *L'incroyable destinée*. Montréal : La courte échelle (collection Roman +), 154 p.

Lavigne, Guy. 1997. *La foire aux fauves*. Montréal: La courte échelle (collection Roman +), 147 p.

Poupart, Jean-Marie. 1992. *Des pianos qui s'envolent*. Montréal : La courte échelle (collection Jeunesse), 93 p.

ANNEXE A : FICHE DE RENSEIGNEMENTS

Bonjour! Merci de répondre à mes questions!

Nom : _____ Fille ☐ Garçon ☐

Enseignant: _____

Année : 3e ☐ 4e ☐ 5e ☐ 6e ☐

Date de naissance : /_____/_____/_____
(jour) (mois) (année)

Quelle langue parles-tu le plus souvent avec tes parents?

Français ☐ Autre langue ☐

Quelle langue parles-tu le plus souvent avec tes frères et tes sœurs?

Français ☐ Autre langue ☐ Je n'ai pas de frère ni de sœur ☐

Quelle langue parles-tu le plus souvent avec tes amis à l'extérieur de l'école?

Français ☐ Autre langue ☐

En quelle année as-tu commencé dans une **classe régulière en français**?

Maternelle	
1 ^{re} année	
2 ^e année	
3 ^e année	
4 ^e année	
5 ^e année	
6 ^e année	

ANNEXE B : TEST EXPRESSIONS FIGÉES

Nom de l'élève: _____

Langue maternelle: _____

Enseignant: _____

Année: 3^e ☐ 4^e ☐ 5^e ☐ 6^e ☐

Consigne : Lis bien les phrases suivantes et complète les mots incomplets sur les tirets.

Exemples :

A) Plus le printemps avance, plus la température augmente. Il fait de plus en p_____ chaud. (Réponse: de plus en plus)

B) D'une journée à l'autre, la température change rapidement. Il faut s'habiller en fonc_____ du temps qu'il fait.

1. En 2008, la ville de Québec fêtera ses 400 ans. De nombreuses activités seront organisées à l'occa_____ de cet événement historique.

2. David a proposé d'organiser une fête pour souligner la fin de l'année scolaire: tout le mon_____ était d'accord.

3. Pour être en forme, nous avons besoin de plusieurs heures de sommeil par nuit. Généralement, un adulte a besoin de p_____ ou m_____ 8 heures de sommeil.

4. De plus en plus de gens se soucient de la qualité de l'environnement.

Beaucoup d'entre eux font des efforts en ce qui concerne le recyclage et l'économie de l'eau.

5. Les prochains Jeux Olympiques auront lieu à Pékin en 2008. En vue de cette grande compétition, les athlètes doivent s'entraîner plusieurs heures par semaine pour être à la hauteur.

6. Autrefois, les grands explorateurs européens n'étaient pas convaincus que la terre était ronde. C'est l'explorateur Magellan qui a prouvé que la notre planète était bel et bien ronde.

7. En voyage, certains vacanciers aiment bien tenir un journal de bord. Ainsi, ils notent leurs découvertes tout au long de leur voyage.

8. Au printemps et en été, les arbres libèrent leur pollen. Pendant cette période, beaucoup de gens sont aux prises avec des allergies saisonnières.

9. Ce que nous apprenons à l'école n'est pas toujours simple. C'est pourquoi le professeur doit bien expliquer de manière à ce que tous les élèves comprennent.

10. À bord d'un avion en pleine altitude, nous ne voyons que des nuages. Puis, au fur et à mesure que nous descendons, nous voyons de plus en plus les voitures et les maisons.

11. En classe, les élèves ne peuvent pas parler tous en même temps. S'ils veulent poser une question, ils lèvent la main et parlent à tour de rôle.

12. Brigitte s'est levée en retard ce matin. Elle s'est dépêchée de se préparer et est sortie à tou_____ vi_____ pour attraper l'autobus.

13. Un an après mon déménagement, j'ai revu mon ami Pierre-Marc. Je ne l'avais pas vu depuis bel_____ lu_____.

14. Parfois, lors d'un orage, nous risquons de manquer d'électricité. C'est pourquoi il est préférable d'avoir une lampe de poche à por_____ de la m_____.

15. Le courriel est un outil de communication de plus en plus utilisé. Il nous permet de communiquer avec des gens provenant des qua_____ coins du m_____.

16. L'enfant porté disparu la semaine dernière a été retrouvé. Heureusement, il est sain et s_____.

17. Ce sera bientôt l'anniversaire de Sophie. Elle rêve de recevoir un beau vélo fl_____ neuf.

18. Dans certaines régions du Québec, les paysages sont magnifiques. En Gaspésie par exemple, nous pouvons apercevoir des montagnes à per_____ de v_____.

Merci pour ta participation!

ANNEXE C : PROTOCOLE DE PASSATION

Protocole de passation

Temps prévu pour l'ensemble du test : 30 minutes

1. L'expérimentateur distribue le test aux élèves.
2. Il attire l'attention des élèves sur la fiche de renseignements (première page). Il demande aux élèves de bien lire les questions jusqu'au bout et d'y répondre du mieux qu'ils peuvent. Il répond aux questions des élèves.
3. Ensuite, l'expérimentateur demande aux élèves de tourner la page. Puis, il lit la consigne du test de locutions à voix haute.
4. Il lit ensuite le premier exemple à voix haute et donne la bonne réponse. Puis, il lit le deuxième exemple à voix haute, mais cette fois-ci les élèves sont invités à donner la réponse oralement. (Les deux exemples sont fournis sur le test des élèves.)
5. L'expérimentateur s'assure que les élèves ont compris la tâche et répond aux questions s'il y a lieu. Cependant, **il ne répond pas aux questions portant sur les locutions recherchées.**
6. À la fin du test, l'expérimentateur récupère les tests complétés.

Consignes particulières

- Le test doit de faire de façon individuelle.
- Aucun ouvrage de référence n'est permis.
- Aucune communication entre les élèves n'est permise.

ANNEXE D : SONDAGE ENSEIGNANTS

Mini-sondage destiné à l'enseignant titulaire

Nom de l'enseignant: _____

Groupe : _____

Année scolaire : _____

École : _____

Au cours de l'année scolaire, avez-vous enseigné des locutions¹ à vos élèves?

Non, jamais ☐

Oui, de façon informelle (sans exercices ni évaluations) ☐

Oui, de façon formelle (avec exercices et/ou évaluations) ☐

Commentaires _____

Merci de votre collaboration!

¹ Exemples de locutions :

nominales : les us et coutumes, coup de foudre, lune de miel

adjectives : bon marché, de rechange, capable de

adverbiales : à la folie, à toute vitesse, à tort et à travers

verbales : jeter de la poudre aux yeux, battre son plein

prépositionnelles : au bout de, à côté de, à destination de, en cas de

discursives : autrement dit, étant donné que, d'autant plus que

propositionnelles (locutions-phrases) : ex : Sauve qui peut!, À table!, Au jeu!, Au secours!.

ANNEXE E : SONDAGE ÉLÈVES

Mini-sondage

Comment as-tu trouvé le test? Entoure ta réponse.

Très difficile plutôt difficile plutôt facile
très facile

Merci pour ta participation!

ANNEXE F : TERMINOLOGIE DU LANGAGE FIGÉ

Source:

Wray, Alison. 2002. *Formulaic Language and the Lexicon*. Cambridge, Angleterre : Cambridge University Press, p. 9.

Amalgams, automatic, chunks, clichés, co-ordinate constructions, collocations, complex lexemes, composites, conventionalized forms, F[ixed] E[xpressions] including I [dioms], fixed expressions, formulaic language, formulaic speech, formulas/formulae, fossilized forms, frozen metaphors, frozen phrases, gambits, gestalt, holistic, holophrases, idiomatic, idioms, irregular, lexical simplex, lexical(ized) phrases, lexicalized sentence stems, listemes, multiword items/units, multiword lexical phenomena, noncompositional, noncomputational, nonproductive, nonpropositional, petrifications, phrases, praxons, preassembled speech, precoded conventionalized routines, prefabricated routines and patterns, ready-made expressions, ready-made utterances, recurring utterances, rote, routine formulae, schemata, semipreconstructed phrases that constitute single choices, sentence builders, set phrases, stable and familiar expressions with specialized subsenses, stereotyped phrases, stereotypes, stock utterances, synthetic, unanalyzed chunks of speech, unanalyzed multiword chunks, units.

ANNEXE G: EXPRESSIONS EXTRAITES DE LA LITTÉRATURE JEUNESSE ET DES MANUELS SCOLAIRES

À bien y penser	À l'inverse de	À mesure que	À ton avis
À cause de	À l'occasion (au sens de «parfois»)	À partir de	À tort et à travers
À ce jour	À l'occasion de	À peine	À tour de rôle
À ce point-là	À la course	À perte de vue	À tout jamais
À cette fin	À la demande de	À peu près	À toute allure
À cœur de jour	À la folie	À portée de la main	À toute vitesse
À court de	À la hâte	À présent	À toutes jambes
À court terme	À la hauteur	À propos (au sens de «pertinent»)	À travers champs
À des fins	À la lumière de	À propos de	À un moment donné
A fait son temps	À la nage	À proximité de	À vive allure
À force de	À la portée de tous	À quel point	À vol d'oiseau
À jamais	À la queue leu leu	À ses yeux	À votre avis
À juste titre	À la suite de	À son tour	Afin de
À l'aide de	À leur convenance	À ta guise	Afin que
À l'aube du 20 ^e siècle	À leur tour	À table!	Ainsi que
À l'échelle de	À long terme	À tel point que	Aller voir ailleurs si j'y suis
À l'égard de	À merveille	À titre d'exemple	Au besoin

Au bord de	Au plus fort de	Aux quatre coins du monde	C'est ainsi que
Au bout du fil	Au plus haut point		C'est peine perdue
Au cœur de	Au point de vue	Avec le recul	C'est pourquoi
Au contraire		Avec le temps	C'est très bien ainsi
Au cours de	Au profit de	Avoir deux poids deux mesures	C'est-à-dire
Au détriment de	Au risque de		Ça m'est égal
Au fil de	Au sein de	Avoir du mal à	
Au fil du temps	Au sujet de	Avoir l'intention de	Ça parle au diable
Au fur et à mesure (que)	Au-dessus de	Avoir recours à	Ce n'est pas tout
Au gré de	Auprès de	Avoir son voyage	Ce n'est pas une mince affaire
Au jeu!	Autant que possible	Bâiller à s'en décrocher la mâchoire	Cette fois-ci
Au jour le jour	Autrement dit		
Au lever du jour	Aux dépens de	Battre son plein	Chercher une aiguille dans une botte de foin
Au lieu de	Aux mains de	Bel et bien	
Au même titre que	Aux milles trouvailles	Bien que	Chose certaine
Au moment de	Aux petits oignons	Boucher un coin	Chose promise, chose due
Au pays des rêves	Aux prises avec	Bras dessus, bras dessous	Comble de malheur

Comme bon te semble	D'un bout à l'autre	Dans quelle mesure	De moins en moins
Comme dans un moulin	D'un côté...d'un autre (côté)	Dans un premier temps	De nos jours
Comme de raison	D'un endroit à l'autre	De cette façon	De nouveau
Comme deux gouttes d'eau	D'une part...d'autre part	De cette manière	De part et d'autre
Comme du monde		De concert avec	De plain-pied
Comme son nom l'indique	Dans la mesure du possible	De façon à	De plus
Comparativement à	Dans la mesure où	De façon à ce que	De plus belle
Compte tenu de	Dans le but de	De façon générale	De plus en plus
Conformément à	Dans le cadre de	De grâce	De sorte que
Côte à côte	Dans le cas de	De haut en bas	De temps à autre
D'ailleurs	Dans le fond	De jour en jour	De temps en temps
D'arrache-pied	Dans les faits	De la part de	De tous bords tous côtés
D'autant plus que	Dans les plus brefs délais	De manière à	De tous les diables
D'autant plus...que...	Dans les règles de l'art	De manière générale	De tout temps
		De même	De toute façon
		De même que	De toutes parts

De toutes pièces	En bonne et due forme	En général	En réalité
De...à...en passant par...	En cas de	En guise de	En revanche
Depuis belle lurette	En ce qui a trait à	En l'honneur de	En somme
Des jours et des jours	En ce qui concerne	En matière de	En suspens
Des kilomètres à la ronde	En ce qui touche	En même temps	En tant que
Dès lors	En ce sens	En mesure de (au sens de «capable»)	En temps et lieu
Différents les uns des autres	En ce temps-là	En outre	En tout temps
Digne de mention	En comparaison de	En parallèle	En un coup de vent
Donner lieu à	En cours de route	En particulier	En vertu de
Du point de vue de	En dépit de	En permanence	En voie de
Du revers de la main	En échange de	En perte de vitesse	En vol plané
Du...jusqu'au ...en passant par...	En effet	En plus	En vue de
En avoir le cœur net	En fait	En principe	Entre autres
	En faveur de	En quel honneur?	Entrer en ligne de compte
	En fin de compte	En quelque sorte	Et ce n'est pas tout
	En fonction de	En raison de	Et voilà que
			Être au courant

Face à face	Glisser un mot sur	La face cachée de la lune	Lune de miel
Faire appel à	Grâce à	La plupart des	Ma foi
Faire en sorte que	Grimper dans les rideaux	La plupart du temps	Mais il en va autrement pour...
Faire le point sur	Histoire de	La vie de tous les jours	Mal en point
Faire rage	Il était une fois	Le long de	Malgré tout
Faire souffler un vent de nouveauté	Il y a de l'électricité dans l'air	Le moins possible	Marché conclu
Faire volte-face	Jeter un coup d'œil	Le moment est venu de	Marquer un point
Fait en sorte que	Joindre les deux bouts	Le plus de...possible	Mener à bien
Faut le faire!	Journal de bord	Les unes les autres	Merveilles du monde
Faute de	Jusqu'à ce que	Les unes sur les autres	Mettre au point
Finir par (+ verbe à l'infinitif)	Jusqu'à présent	Les uns avec les autres	Mettre en évidence
Flambant neuf	L'avocat du diable	Les uns envers les autres	Mettre en œuvre
Font appel à	L'un à côté de l'autre	Leurs cliques et leurs claques	Mettre fin à
Gagner du temps	L'un après l'autre	Lors de	Mettre la dernière main à
			Mine de rien
			Mis à part

Mis sur pied	Par la même occasion	Perdre les pédales	Quoi qu'il en soit
Mode de vie	Par la suite	Petit à petit	Ravalier sa salive
N'ont pas de secret pour lui	Par le fait même	Peu à peu	Regarder de travers
N'ont qu'à bien se tenir	Par rapport à	Peu importe	Relativement à
Ne pas avoir froid aux yeux	Par-ci par-là	Pleurer à chaudes larmes	Remettre en question
Ne pas déborder de	Parler à travers son chapeau	Plus ou moins	Risquer le tout pour le tout
Ne pas voir plus loin que le bout de son nez	Partir en peur	Point de vue	Rôle de premier plan
Non seulement...mais (aussi)...	Pas de tout repos	Poudre aux yeux	S'en donner à cœur joie
Par ailleurs	Pas sorti du bois	Pour ce faire	S'il y a lieu
Par conséquent	Pas un mot plus haut que l'autre	Pour ma part	Sain et sauf
Par contre	Passer sous silence	Prendre à cœur	Sans aucun doute
Par exemple	Prendre en main	Prendre les jambes à son cou	Sans cesse
Par l'intermédiaire de	Penser tout haut ce que les autres pensent tout bas	Prendre les	Sans doute
Par la force du nombre	Perdre le fil	Quant à	Sans plus tarder
		Que le diable l'emporte	Sans précédent
		Qui sait	

Sans relâche	Tandis que	Tout à fait	Us et coutumes
Sans se faire prier	Tant bien que mal	Tout à tour	Vendre son âme au diable
Sauve qui peut!	Tant et aussi longtemps que	Tout au long de	Venir à bout de
Se confondre en excuses	Tant et si bien que	Tout laisse croire que	Vice versa
Se mettre à l'abri	Tantôt...tantôt ...	Tout le monde	Vis-à-vis (de)
Se mettre les yeux en face des trous	Tels quels	Toute la misère du monde	Voir le jour
Si jamais	Tête-à-tête	Toutes sortes de	
Sont autant de...qui...	Tirer le diable par la queue	Trembler de peur	
Sont autant d'exemples de...	Tirer parti de	Un coup d'œil	
Sous forme de	Tirer profit de	Un de ces jours	
Subvenir à leurs besoins	Tomber à point	Une à côté de l'autre	
Sur le plan (de)	Tour à tour	Une à la fois	
Sur un coup de tête	Tour du monde	Une chose est sûre	
Sur une base volontaire	Tourner le fer dans la plaie	Une foule de	
	Tous plus... les uns que les autres	Une image vaut mille mots	